



Szene Z4

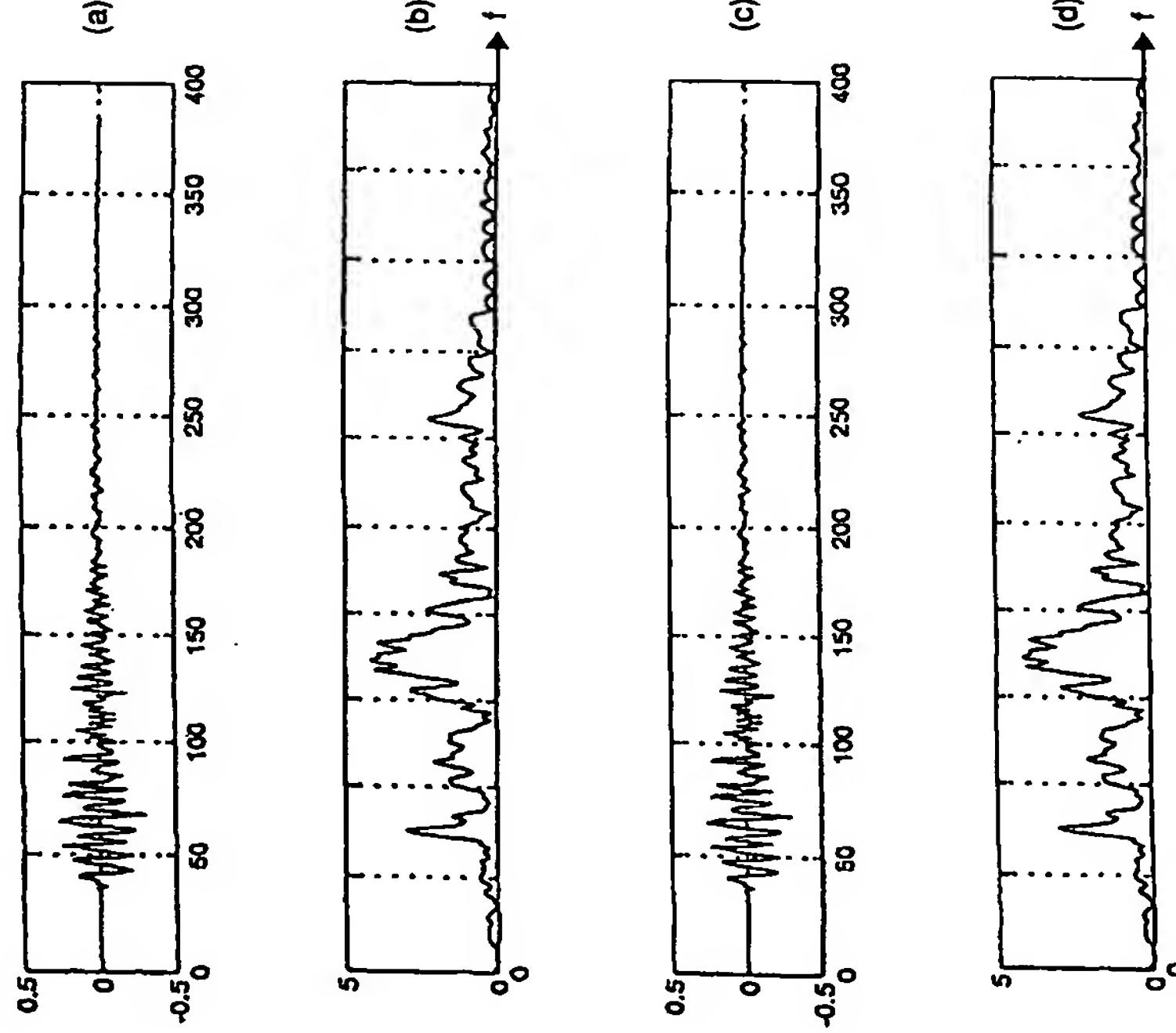


FIG.28



EUROPÄISCHE PATENTANMELDUNG

(12)

(43) Veröffentlichungstag:

06.10.1999 Patentblatt 1999/40

(21) Anmeldenummer: 99106123.5

(22) Anmeldetag: 01.04.1999

(11)

(51) Int. Cl. 6: H04R 3/00

(54) Verfahren zur Befreiung eines Mikrophonsignals

(84) Benannte Vertragsstaaten:
AT BE CH CY DE DK ES FI FR GB GRI IT LU
MC NL PT SE
Benannte Erstreckungsstaaten:
AL LT LV MK RO SI

(30) Priorität: 03.04.1998 DE 19814971

(71) Anmelder:
DaimlerChrysler Aerospace AG
81663 München (DE)
(72) Erfinder: Thomas, Hans-Jürg
89233 Neu-Ulm (DE)

(55) Verfahren zur Störbefreiung eines Mikrophonsignals
(57) Für ein Verfahren zur Befreiung eines Mikrophonsignals von Störsignalanteilen durch Erzeugen eines Kompressionssignals und Subtraktion des Kompressionssignals vom Mikrophonsignal wird vorgeschlagen, die Komensation vollständig im Frequenzbereich vorzunehmen und auch das Ausgangssignal im Frequenzbereich zu verarbeiten. Es werden Maßnahmen zur Verringerung des Aufwands bei der Signaverarbeitung angegeben. Vorteilhafte Weiterbildungen sehen beispielsweise vor, daß zur Störbefreiung eines Sprachsignals eine Filtereinstellung verwendet wird, welche in einer vorhergehenden Sprachpause gewonnen wurde, und zwar das für lange Impulsantworten das NachbildungsfILTER in mehrere Filter aufgespalten wird. Die Erfindung ist insbesondere vorteilhaft für die Befreiung eines Spracheingabesignals in einem Fahrzeug von Störsignalanteilen eines Radios oder dgl., dessen Quellsignal als Referenzsignal zur Verfügung steht.

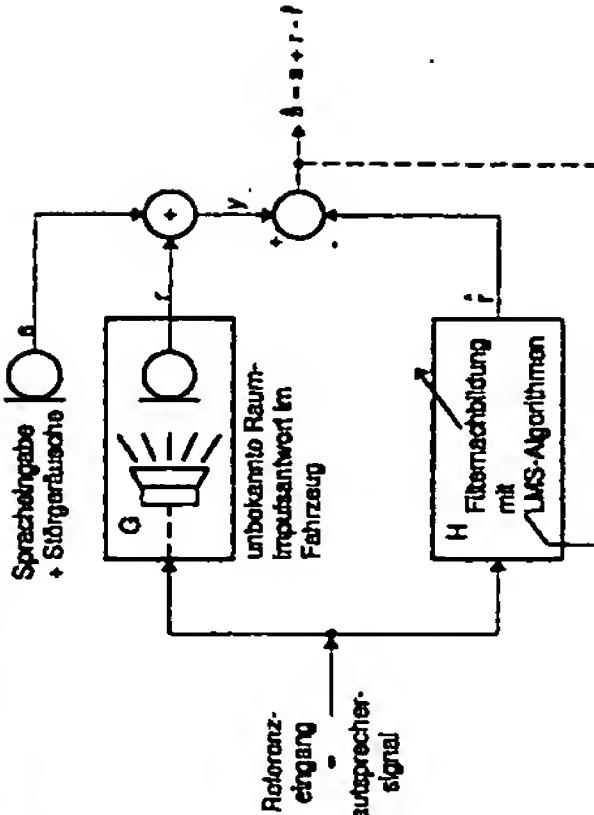


FIG.1

EP 0 948 237 A2

BEST AVAILABLE COPY

Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Störfehlung eines Mikrophonsignals.

[0002] Derartige Verfahren gewinnen insbesondere für die Spracheingabe von Kommandos und/oder für Freisprechtelefonie zunehmend an Bedeutung, wobei insbesondere die Situation in einem Fahrzeug ein wichtiger Anwendungsfall ist.

[0003] Eine besondere Situation ist in Fahrzeugen häufig dadurch gegeben, daß ein Wiedergabegerät wie z.B. ein Radio, ein Kassetten- oder CD-Abspieler über einen Lautsprecher eine Geräuschungsgabe erzeugt, die als Störsignal ein vom Mikrofon aufgenommenes Sprachsignal, beispielsweise für die Spracherkennung oder Telefonübertragung überlagert. Zur Erkennung von Spracheingaben in einem Sprachkennner oder zur verständlichen Sprachübertragung über Telefon ist das Mikrophonsignal soweit wie möglich von Störsignalanteilen zu befreien.

[0004] Das von einer Störquelle, insbesondere einem Lautsprecher ausgehende Störsignal gelangt nicht nur auf direktem kürzestem Weg zum Mikrofon, sondern tritt auch noch über zahlreiche Reflexionen auf. Die gesamte Einwirkung des Störsignals von der Störquelle auf das Mikrophonsignal kann durch eine a priori unbekannte Übertragungsfunktion des Raumes, bei spielsweise des Fahrgastrumes eines Kraftfahrzeugs beschrieben werden. Die Übertragungsfunktion ändert sich je nach Besetzung des Fahrzeugs und nach Position der einzelnen Personen. Durch Nachbildung dieser Übertragungsfunktion und Führung eines Referenzsignals von der Störquelle mit dieser Nachbildung kann ein Kompressionsignal erzeugt werden, welches durch Subtraktion vom Mikrophonsignal ein vom Störsignal befreites Signal, beispielsweise ein reines Sprachsignal liefert. Im Realfall stellt die genannte Nachbildung eine mehr oder minder gute Annäherung an die unbekannte Übertragungsfunktion dar und die Störung kann nicht vollständig beseitigt werden.

[0005] Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zur Störfehlung eines Mikrophonsignals anzugeben, daß bei vertretbarem Signalerarbeitungsaufwand gute Eigenschaften hinsichtlich der Entstörung aufweist.

[0006] Die Erfindung ist im Patentanspruch 1 beschrieben. Die Unteransprüche enthalten vorbehaltete Ausgestaltungen und Weiterbildungen der Erfindung.

[0007] Wesentlich an dem erfindungsgemäßen Verfahren ist, daß die Kompression des Störsignalanteils im Mikrophonsignal mittels eines aus dem Referenzsignal über die Nachbildung der Übertragungsfunktion erzeugten Kompressionssignals im Frequenzbereich vorgenommen wird, so daß Mikrophonsignal, Kompressionssignal und Ausgangssignal im Frequenzbereich d.h. in Form von Spektren vorliegen. Die Signaverar-

beitung in diesem Verfahrensschritt im Frequenzbereich erfordert zwar eine spezielle Transformation des Mikrophonsignals, berücksichtigt aber, daß die Nachbildung der Übertragungsfunktion im Frequenzbereich vorteilhafter ist und stellt für eine vorteilhafte nachfolgende zusätzliche Geräuschreduktion des Ausgangssignals, die typischerweise gleichfalls im Frequenzbereich vorgenommen wird, bereits eine besonders geeignete Signalform bereit.

[0008] Durch einfache Näherungen beim Ersatz eines Verarbeitungsschritts mit einem Zeitfenster kann durch Übergang zu einer Faltung im Frequenzbereich eine deutliche Reduzierung des Verarbeitungsaufwands realisiert werden.

[0009] Für lange Impulsenantworten der Übertragungsfunktion bzw. deren Nachbildung sieht eine vorteilhafte Weiterbildung der Erfindung eine Aufteilung des Nachbildungsfilters in mehrere Teiltüller zu zehn verschiedenen Segmente des segmentierten Referenzsignals vor, deren Koeffizienten-Aktualisierung zeitlich gestaffelt sein kann, wodurch der Signaverarbeitungsaufwand gering gehalten werden kann.

[0010] Als besonders vorteilhaft erweist es sich, die Ersttönung eines Sprachsignals auf der Basis einer Einstellung des Nachbildungsfilters, die in einer vorhergehenden Sprachpause gewonnen und gespeichert wurde, vorzunehmen.

[0011] Die Aufteilung des Nachbildungsfilters in mehrere Teiltüller und die Störbereitung auf der Basis einer in einer Sprachpause gewonnenen Filtereinstellung sind auch unabhängig von der Störsignalkompenstation im Frequenzbereich eigenständig für die Störberichtigung eines Mikrophonsignals realisierbar und vorteilhaft.

[0012] Die Erfindung ist nachfolgend anhand von bevorzugten Ausführungsbeispielen unter Bezugnahme auf die Abbildungen noch eingehend veranschaulicht. Dabei zeigt:

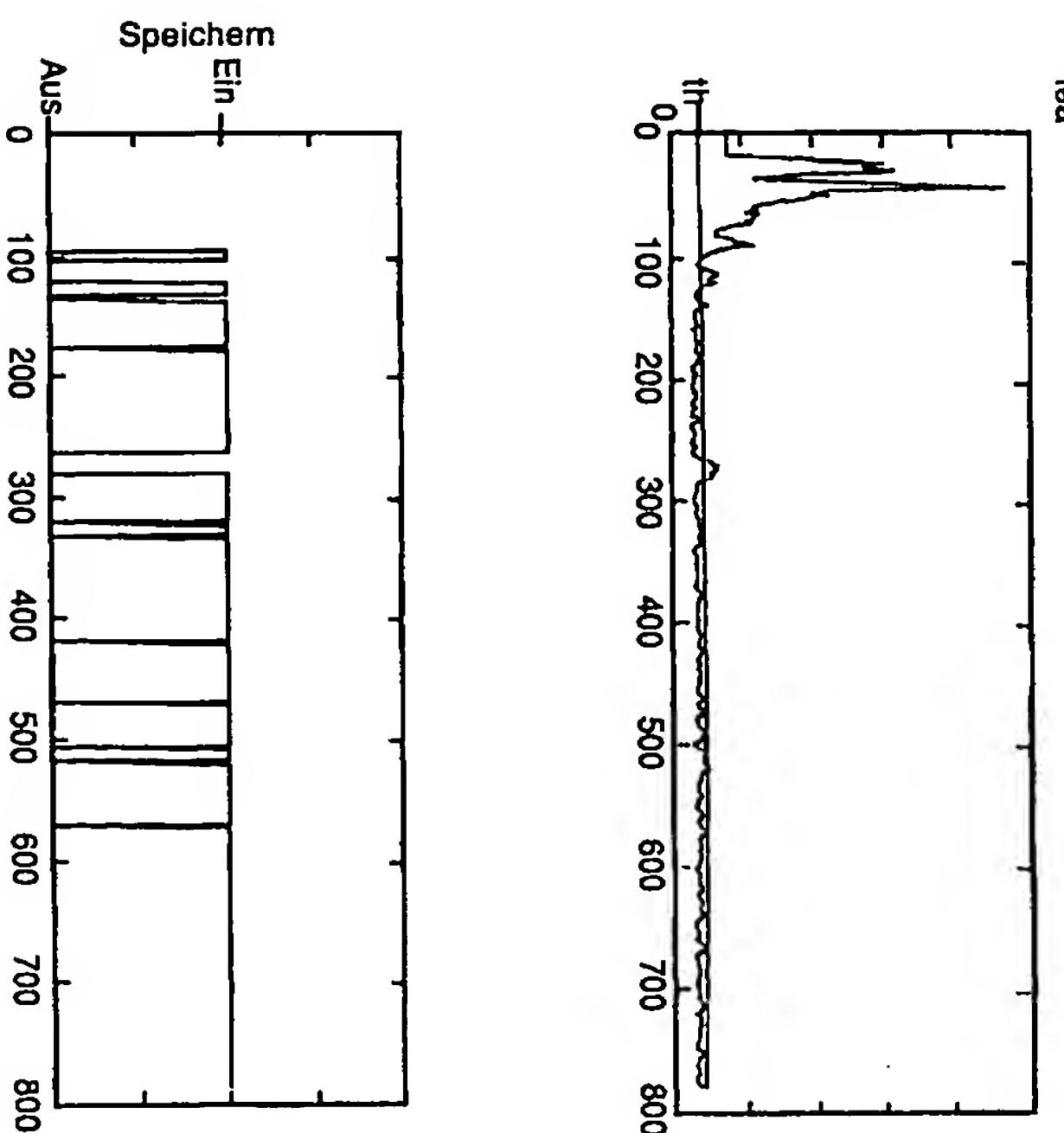


FIG.27

Szene Z4

fea

der Filtereinstellungen

Fig. 9 Signale einer synthetischen Beispielszene

Fig. 10 Impulsantwort und Übertragungsfunktion zu Fig. 9

Fig. 11 Signal einer ersten Meßszene

Fig. 12 Impulsantwort und Übertragungsfunktion zu Fig. 11

Fig. 13 das Beispiel nach Fig. 11 mit Speicherung der Filtereinstellungen

Fig. 14 eine Sprachpausendektion zu Fig. 13

Fig. 15 Impulsantworten und Übertragungsfunktion zu Fig. 11 und Fig. 13

Fig. 16 Übergang von einem Zeitfenster zu einer Faltung im Frequenzbereich

Fig. 17 ein Rechteck-Zeitfenster mit Unienspektrum

Fig. 18 ein Hamming-Zeitfenster mit Unienspektrum

Fig. 19 Staffelung von Signalblöcken bei der Filterberechnung

Fig. 20 Signale einer zweiten Meßszene

Fig. 21 eine Sprachpausendektion zu Fig. 20

Fig. 22 Impulsantworten und Übertragungsfunktion zu Fig. 20 und Fig. 21

Fig. 23 Signale einer dritten Meßszene

Fig. 24 eine Sprachpausendektion zu Fig. 23

Fig. 25 Impulsantworten und Übertragungsfunktion zu Fig. 23 und Fig. 24

Fig. 26 Signale einer vierten Meßszene

Fig. 27 eine Sprachpausendektion zu Fig. 26

Fig. 28 Impulsantworten und Übertragungsfunktion zu Fig. 26 und Fig. 27

Spracheingabesystems. Unter der Annahme, daß sich die Übertragungsstrecke G dann nach als Transversalfilter mit einer gewöhnlichen Summe zeitlich verzögter Echos darstellt, läßt sich eine Filternachbildung H finden, die im Idealfall $H \cdot G$ eine vollständige Kompensation des Radiosignals ermöglicht.

[0014] Das Lautsprechersignal x wird durch die a priori unbekannte Übertragungsfunktion G des Fahrzeugraumraums gefiltert. Es entsteht die Störkomponente r , die sich mit dem Sprachsignal s zu dem Mikrophonsignal y addiert. Um die Störkomponente r zu kompensieren, wird mittels der Filternachbildung H ein Schätzwert r^* aus dem Lautsprechersignal x erzeugt. Der Ausgang der Schaltung ist der Schätzwert für das Sprachsignal:

$$s^* = s + r \cdot r^* = s + E$$

[0015] Dem Sprachsignal s ist also am Ausgang der Schaltung noch das Fehlersignal $E = r^*$ überlagert, welches in der Praxis möglichst klein gehalten werden sollte. Das Sprachsignal kann noch Störungen in Form von z.B. Motorgeräuschen oder externen Geräuschen enthalten, die aber in diesem Zusammenhang nicht explizit behandelt werden.

[0016] H ist ein adaptives Filter und arbeitet nach einem in der Literatur bekannten Standardverfahren, dem LMS-Algorithmus (least mean squares). Neben dem Eingangssignal x wird noch das Fehlersignal E benötigt, um die Koeffizientenadaption im Filter H zu bewerkstelligen. Hierfür ist das Ausgangssignal s^* der Bestimmung der Filterkoeffizienten zugeführt.

[0017] Fig. 2a zeigt in anderer Darstellung nochmals die Anordnung von Fig. 1 als Radiosignalkompenstation. Das adaptive System H kann z.B. im Zeitbereich als FIR-FILTER (finite-impulse-response-FILTER) realisiert werden. Bei großen Impulsantwortlängen, wie sie in der Praxis häufig auftreten, ist hierzu allerdings ein sehr hoher Rechneraufwand notwendig. Verschiedene Varianten gegenüber einer Zeitbereichslösung bietet die Realisierung des LMS-Algorithmus im Frequenzbereich (FLMS). Wegen der blockweisen Verarbeitung von Daten in den als diskreten Fouriertransformationen realisierten spektralen Transformationen und der Filterrealisierung im Frequenzbereich durch Multiplikationen wird dieses Verfahren besonders recheneffizient.

[0018] Fig. 2b zeigt ein Blockschaltbild des FLMS-Algorithmus. Die zugehörige Theorie ist an sich bekannt und daher an dieser Stelle nicht im Detail behandelt. Es bedeuten F eine spektrale Transformation FFT eines Zeitsignals in den Frequenzbereich und F^{-1} die inverse FFT. Die als Projektionen P_1 , P_2 und P_3 bezeichneten Verarbeitungsschritte dienen der korrekten Segmentationierung der Daten durch die blockweise Verwendung mit der FFT bzw. IFFT und werden später noch genauer erläutert. Die Arbeitsweise des Filters besteht in der Multiplikation des Referenzspektrums X mit dem Filter-Koeffizientenvektor H . Das Spektrum des Lautsprechers abgestrahlte akustische Signal gelangt auf direktem Wege, aber auch über zahlreiche Reflexionen im Fahrzeugraumraum, auf das Mikrofon des

Szene Z4

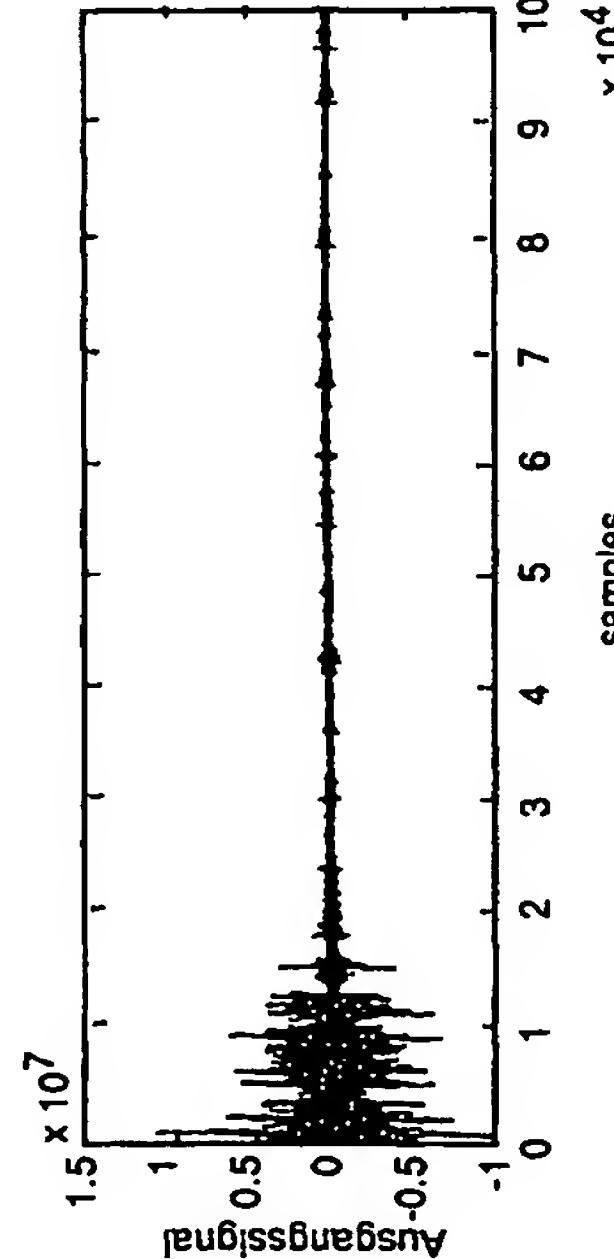
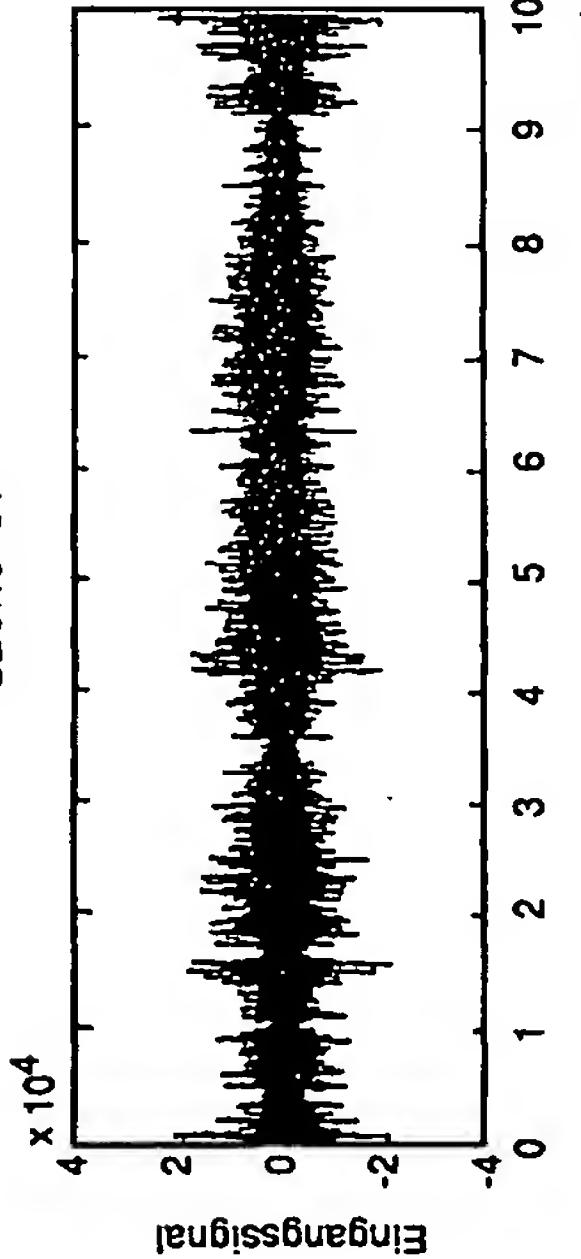


FIG.26

Filterausgangs R^a wird über F^{-1} zurück in den Zeitbereich transformiert. Nach Anwendung der Projektion P_2 auf den Realteil des so erhaltenen Kompressionssignals steht das Signal r^a zur Verfügung. Die Differenz der Signale

$$s^a = y - r^a = s + r - r^a = s + E$$

stellt den eigentlichen Ausgang, eine Schätzung der Spracheingabe, dar.

[0019] Wesentlicher Bestandteil des adaptiven Filters ist die Koeffizientenadaption im Block K, die im Fig. 2b durch die Erneuerungsgleichung

$$H' = H + \Delta H'$$

beschrieben wird. Die hier mit zwei speziellen Transformationen besonders aufwendige Projektion P_1 berechnet aus H' den für die Filterung benötigten Koeffizientenvektor H . Zur Berechnung des Korrekturvektors $\Delta H'$ wird neben dem Referenzspektrum X das Spektrum S^a des mit P_3 berechneten Ausgangssignals $s + r - r^a$ benötigt.

[0020] Ein detailliertes Blockschaltbild des in Fig. 2b dargestellten FLMS-Algorithmus zeigt Fig. 3. Die Abtastwerte eines Signals und die Stützstellen der FFT seien in gebräuchlicher Weise als samples bezeichnet. Alle Spektraltransformationen und deren Inverse sind als 256-Punkte - FFTs, die jeweils um 128 samples überlappen, zu segmentieren. Zu beachten ist, daß sich das Ausgangssignal s^a im Zeitbereich aus 128-sample-Blöcken zusammensetzt. Es entsteht aus der Differenz der zweiten Blockhälften (also jeweils der samples 129 bis 256) von Mikrofonsignal und gefiltertem Kompressionsignal r^a . Aufwendig ist die Projektion P_1 , die 256 FFTs benötigt und den Vektor H' in den Vektor H umrechnet. Hilfsel wird aus dem komplexen 256-Punkte-Ergebnisvektor der Rücktransformation vom

Frequenz- in den Zeitbereich (IFFT) die erste Hälfte (samples 1 bis 128) ausgeschnitten und die zweite Hälfte (samples 129 bis 256) zu Null gesetzt. Nach Anwendung dieses Rechteckkerns im Zeitbereich erfolgt wieder mittels FFT die Transformation in den Frequenzbereich. Einfach ist die Projektion P_2 . Sie besteht aus der oben schon beschriebenen Ausschnittsbildung der letzten 128 samples, wodurch aus überlappenden 256-sample-Blöcken wieder nicht überlappende 128-sample-Blöcke entstehen. Ebenfalls sehr einfach ist schließlich auch die Projektion P_3 , welche umgekehrt aus nicht überlappenden 128-sample-Blöcken des Ausgangssignals durch Vorausstellen von 128 Nullwerten wiederum überlappende 256-sample-Blöcke bereitstellt. Die Adaption der Filterkoeffizienten H_{L+1}^a für einen Zyklus $L+1$ besteht aus der Addition eines Erneuerungsvektors $\Delta H'_L$ zum alten Koeffizientenvektor H_L^a . Diese Erneuerung errechnet sich aus dem Produkt zwischen dem Spektrum S_L^a des Ausgangssignals und dem konjugiert komplexen Spektrum X_L^a des Referenz-

signals - gewichtet mit einer speziellen Leistungsnormierung $2\mu_L$, $\Delta H'_L = 2\mu_L \cdot X_L^a \cdot S_L^a$. Zum Zweck dieser Leistungsnormierung ist der mit einer Konstanten 2α multiplizierte Kennwert des geglätteten Referenzspektrums $S_{x,L}$ zu berechnen

$$2\mu_L = 2\alpha S_{x,L}, \text{ wozu ein rekursives Filter 1. Ordnung mit einer Konstanten } \beta \text{ dient}$$

$$S_{x,L} = \beta \cdot |X_L|^2 + (1-\beta) \cdot S_{x,L-1}$$

[0021] Die Arbeitsweise des LMS-Algorithmus wird erheblich von der Adaptionkonstante α und der Gittungskonstante β beeinflußt. Zwischenstufen in Rekursioschleifen sind mit Sp bezeichnet.

[0022] Die bisher beschriebene Anordnung des FLMS-Algorithmus erlaubt Filtermischbildungen mit einer maximalen Impulsantwortlänge von einer halben FFT-Länge, im Beispielfall also 128 samples. Sollen längere Impulsantworten kompensiert werden, ist der schon bekannte LMS-Algorithmus für einen Teilfilter (Fig. 4a) auf n Teilfilter zu erweitern. Eine 3-Teilfilter-Lösung mit einer Impulsantwortlänge von $3 \cdot 128 = 384$ samples hat sich bei der Radiosignalumwandlung im PKW mit einem Spracheingabesystem bewährt (Fig. 4b). Der im Fig. 4a mit B bezeichnete Block mit den Eingangssignalen X und S^a und dem Kompressions-Spektrum R^a als Ausgang ist durch die im Fig. 4b dargestellte Erweiterung zu ersetzen. Das Spektrum X des Referenzsignals wird durch Zwischenspeicher D um 1 bzw. 2 Blocklängen verzögert und das unverzögerte X_1 und die beiden verzögerten Spektren X_2, X_3 werden separat in mit in einer erweiterten Projektion P_1 getrennt bestimmten Koeffizientenvektoren H_1, H_2, H_3 multipliziert. Die Bildung der Koeffizientenvektoren erfolgt analog zum Fall nur eines Teilfilters, wobei in K_1, K_2, K_3 jeweils das zugehörige Referenzspektrum mit dem Spektrum S^a des Ausgangssignals verknüpft wird. Der Aufwand wird hauptsächlich durch die Verdrehung der Projektion P_1 beträchtlich erhöht. Zusätzlich ist der Speicherplatzbedarf wird notwendig um die Spektren des um 1 bzw. 2 Blocklängen zeitlich älteren Referenzsignals X bereitzustellen.

[0023] Bei der beispielhaft angegebenen Aufgabenstellung der Unterdrückung des Radiosignals bei Spracheingabe im Kfz ist es vorteilhaft die Ausgangsdaten nicht im Zeit- sondern im Frequenzbereich auszugeben, da dadurch eine verbesserte Anpassung an eine nachgeschaltete Geräuschunterdrückung erreicht werden kann. Der bereits vorgestellte LMS-Algorithmus mit einem Teilfilter benötigt gemäß Fig. 5a insgesamt 5 FFTs bei einem Ausgangssignal im Zeitbereich. Wird dem Ausgang eine FFT nachgeschaltet, erhöht sich der Aufwand bei einem Frequenzbereichs-Ausgangssignal auf 6 FFTs. Die gleiche FFT-Anzahl ergibt sich zunächst auch bei einer äquivalenten Lösung nach Fig. 5b. Diese Variante besitzt jedoch folgende Vorteile:

- Bei der zeitgleichen Spektralanalyse der Signale x

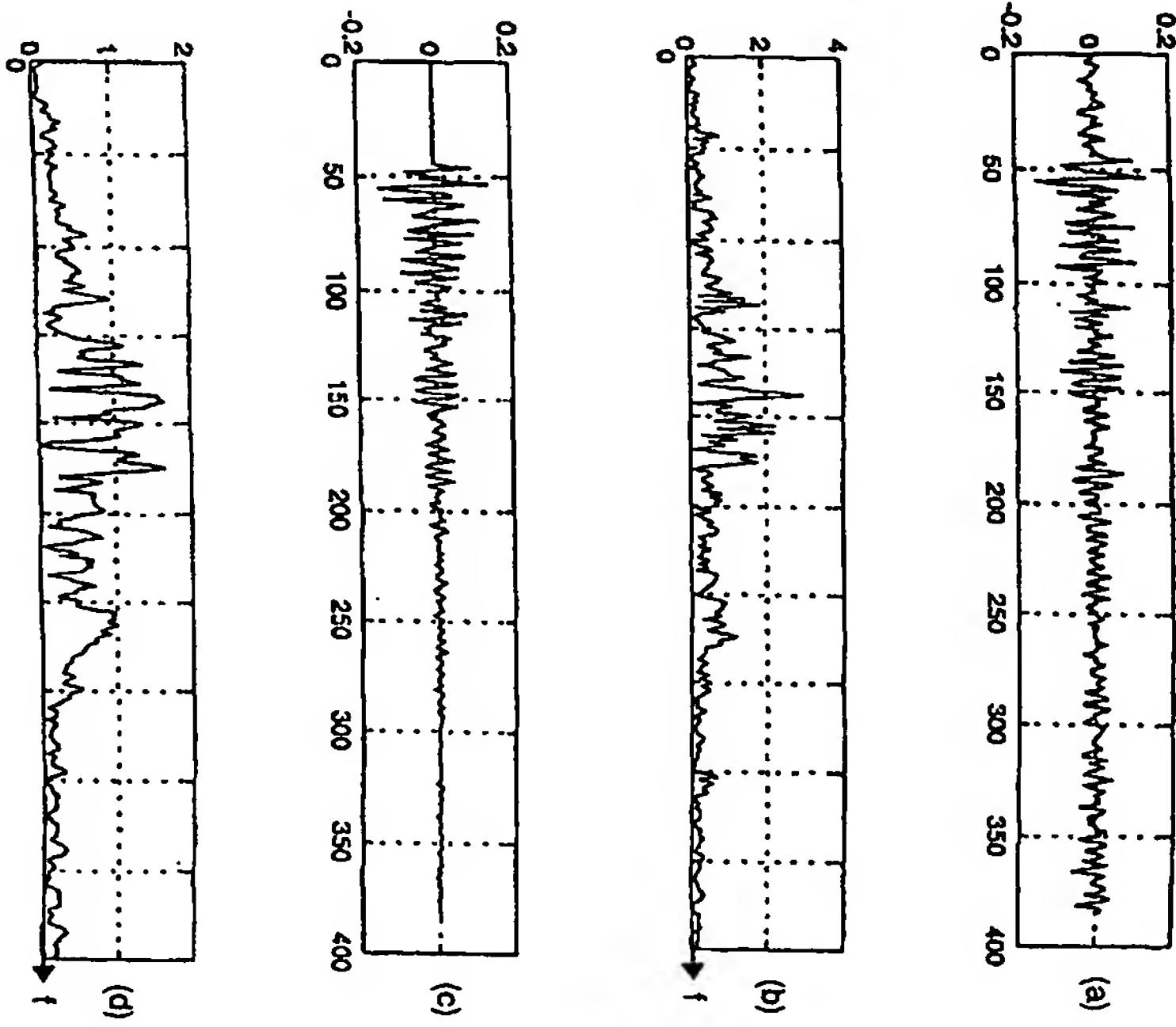


FIG.25

und y ist nur eine einzige 256-Punkte-FFT mit geringem Zusatzaufwand für eine spektrale Separation notwendig. Man erzielt eine Einsparung von 1 FFT.

- Die hier mit P_4 gekennzeichnete und neu definierte Projektion ist bis auf das verwendete Zeitfenster formal identisch mit der Projektion P_1 . Wie später gezeigt wird, läßt sich P_4 durch eine relativ einfache Faltungsoperation im Frequenzbereich erzielen, ohne daß eine merkliche Einbuße an Qualität in Kauf genommen werden muß. Man erzielt eine Einsparung von 2 FFT's.

[0024] Fig. 6 stellt ein ausführlicheres Blockschaltbild des FLMS-Algorithmus mit Frequenzbereichs-Ausgangssignal dar und gestattet nochmals einen Vergleich mit Fig. 3 (Zeitbereichs-Ausgang). Unverändert geblieben ist die Filteradaptation bestehend aus Glättung der spektralen Leistung, Leistungsnormierung und Koeffizientensteuerung. Neu sind die FFT im Mikrophonkanal, die Differenzbildung $Y - R$ im Frequenzbereich zur Ausgangsbildung, und schließlich die neu definierte Projektion P_4 , welche sich lediglich durch das komplementäre Zeitbereichsfenster von der Projektion P_1 unterscheidet.

[0025] Als Vorstufe einer nachfolgend beschriebenen bevorzugten Ausführung betrachte man Fig. 7. Dargestellt ist der FLMS-Algorithmus mit 3-Tellfiltern (384-sample-Impulsantwort), welcher eine ausreichende Unterteilung des Radiosignals im Mikrophonkanal und P_4 sind vereinfacht dargestellt. Es ist der schon aus Fig. 4b bekannte zusätzliche Aufwand in Form der spez. P sowie die Verfeinung der Projektion P_1 und P_4 sind vereinfacht dargestellt. Es ist der schon aus Fig. 4b bekannte zusätzliche Aufwand in Form der spez. P sowie die Verfeinung der Projektion P_1 ersichtlich. Im Gegensatz zur 1-Tellfilter-Lösung nach Fig. 6 wird die Summe W von gegenwärtigem und den beiden zeitlich vorausgegangenen Referenz-Leistungspekten auf den Eingang des rekursiven Filters gegeben. Die Tatsache, daß am Filterausgang nun praktisch die 3-fache geglättete spektrale Leistung vorliegt, wird nach der Kehwertbildung durch Multiplikation mit der Konstanten 6μ berücksichtigt. Nach der spektralen Leistungsnormalierung des in P_4 modifizierten Ausgangsspektrums S^* wird die Filter-Adaption nun für die 3 Koeffizientenvektoren der 3 Tellfilter separat durchgeführt.

[0026] Ein Beispiel 20 für die Wirkungsweise der Erfindung nach Fig. 7 zeigt Fig. 9. Die Eingangsdaten wurden synthetisch generiert. Das Referenzsignal X stellt 100000 Abtastwerte eines weißen Gaußrauschens bei einer Abtastfolgefrequenz von $f_s = 12$ kHz dar. Das Mikrophonsignal Y entstand durch Faltung dieses Rauschsignals mit einer ebenfalls konstruierten 384-sample-Impulsantwort sowie der Addition eines außenstehenden Sprachsignals. Beim Abhören die gesprochenen Ziffern gerade noch im farbigen (weiß gesprochenen) Bereich zu erkennen. Das in den Zeitbereich zurücktransformierte Ausgangssignal des Schalters befindet nach einem ca. 1 Sekunde (12000 samples) dauernden Einschwingvorgang sehr wirkungsvoll die Spracheingabe vom Rauschen und liefert ein unverzerrtes jedoch leicht verhältnisses Sprachsignal S^* (Fig. 9 unten). Die verwendeten beiden Parameter waren $\alpha = 0,05$ und $\beta = 0,5$. Werte, die sich auch bei den später vorgestellten Beispielen gut bewährt haben.

[0027] Aus den jeweils 129 samples langen Tellkoeffizientenvektoren H_1 , H_2 , H_3 der 3 Tellfilter nach Bild 7 läßt sich nun zu jedem Zeitpunkt die hieraus resultierende 31128-sample-Impulsantwort bzw. die zugehörige Filterübertragungsfunktion berechnen. So zeigt Fig. 10 oben die 384-sample-Impulsantwort, wie sie sich ganz am Ende der Szene - also nachdem die Ziffer "0" gesprochen wurde - ergibt. Sie ist ein sehr gerausches Abbild derjenigen Impulsantwort, die zur Faltung mit weißen Gaußrauschen und damit zur synthetischen Generierung des Signales mikro benutzt wurde. Die zugehörige Betriebsübertragungsfunktion (Fig. 10 unten) im Bereich zwischen den Frequenzen 0 und $f_s/2 = 6$ kHz stellt einen mit zahlreichen schmalbandigen Resonanzüberhöhungkeiten behafteten Tiepflätsfrequenzgang dar.

[0028] Weißes Rauschen als Referenzeingangssignal und gefiltertes "farbiges" Rauschen als Mikrophonanfangssignal stellen im Sinne der Aufgabenstellung, eine Nachbildung dieses Filters zu finden, den einfachsten Fall dar. Da das Referenzsignal per Definition alle Frequenzteile enthält, gelingt die Filteradaptation hier am schnellsten. Die zusätzliche additive Spracheingabe im Mikrophonanfangssignal - also das eigentliche Nutzsignal des Spracheingangssignals - stellt für den (FLMS-)Algorithmus eine Störung dar, welche die korrekte Adaptation der Filterkoeffizienten behindert. Anders ausgedrückt: nur in Sprachpausen ist das System dazu in der Lage die Raumakustik des Fahrzeugraums (Strecke Radiolautsprecher zum Mikrofon) richtig nachzubilden und dadurch eine Korrektion der Radiowiedergabe zu bewirken. Im oben demonstrierten Beispiel gemäß Fig. 9 gelingt dies sehr gut, da der Mikrofon-eingang im wesentlichen aus Rauschen und nur zu einem sehr geringen Teil aus Spracheingabe besteht.

[0029] Aus solchen Messungen im Fahrzeug hingegen entstammt das an den Radiolautsprechern dem abgegebene Referenzsignal radio und das vom Mikrofon des Spracheingabesystems aufgezeichnete Signal mikro der Szene 21. Dieses Mikrophonsignal ist in Fig. 11 oben dargestellt, besteht aus 100000 samples und besitzt demnach bei einer Sampling-Frequenz von 12 kHz eine zeitliche Dauer von ca. 8,3 Sekunden. Es handelt sich um flüssig und relativ schnelle gesprochene Sprache eines hinter rechts im Auto sitzenden Fahrgässen, während gleichzeitig Musik mit normaler Lautstärke aus dem Autoradio-Lautsprecher erklingt.

Nach Anwendung der Entstörmaßnahme gemäß Fig. 7

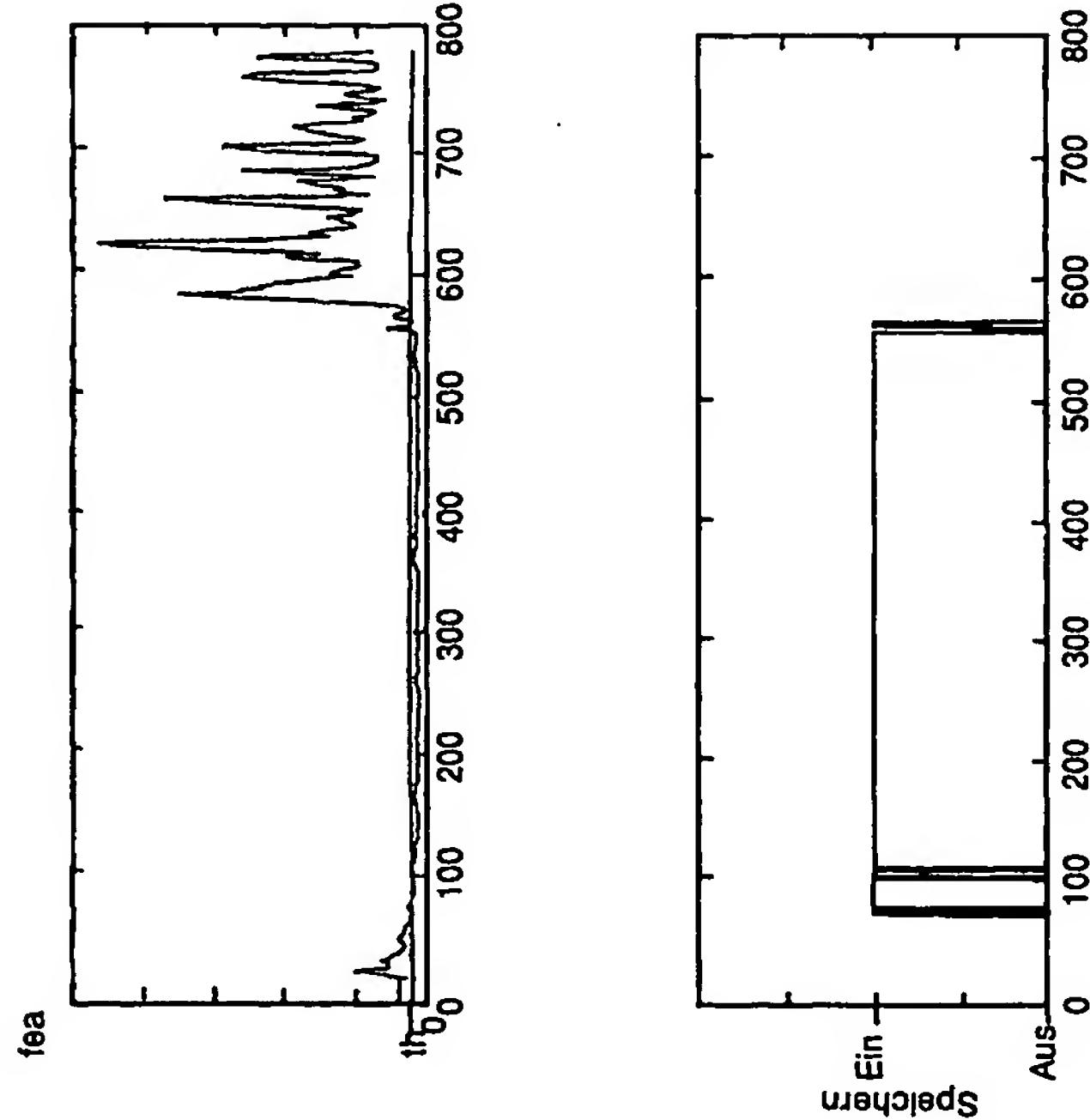


FIG.24

Fig. 11 unten dargestellte Ausgangssignal. Der Hörfehl-Nachteil ist jedoch, daß die erwünschte Radiospaustenberichtigung in starkem Maße davon abhängt, ob gerade gesprochen wird oder nicht. Da wieder am Szenenende ermittelte 384-sample-Impulsantwort mitzugehöriger Übertragungsfunktion ist aus Fig. 12a ersichtlich. Eine korrekte Impulsantwort ist an den typischen Nullsamples (Totzeit) am Anfang zu erkennen, welche von der Laufzeit des Direktschalls vom Radiolautsprecher zum Mikrofon herführen. Aus den hier vorhandenen starken Störungen am Anfang sowie am Ende der Impulsantwort läßt sich demnach der Schluß ziehen, daß die Filteradaptation an dieser Stelle wegen vorhandener Spracheingabe außerst unzureichend ist [10030]. Die im folgenden anhand von Fig. 8 beschriebene Ausführung form beruht auf folgender Grundidee: ein geeignetes Merkmal dient zusammen mit einem Schwellenwert als Indikator für eine Spracheingabe. Unterschreitet das Merkmal die Schwellen, so ist dies ein Anzeichen für fehlende Spracheingabe. In diesem Fall kann - wie oben schon festgestellt - eine weitgehend ungestörte Filteradaptation erfolgen. Bei Spracheingabe wird nun auf denjenigen Filterkoeffizientensatz zurückgegriffen, der unmittelbar vor der Schwellenüberschreitung - d.h. am Ende der vorangegangenen Sprachpause - abgespeichert wurde. Diese gespeicherten Koeffizienten H10, H20, H30 liefern im Regelfall eine deutlich bessere Radiosignal-Kompensation als die unter dem störenden Einfluß der Spracheingabe sich ständig ändernden aktuellen Koeffizienten H, H2 H3.

[10031] Fig. 8 stellt eine Ausführung mit einer weiter verbesserten FLMS-Verarbeitung mit 3 Teilträgern dar. Neben den schon in Fig. 7 vorhandenen aktuellen Filterkoeffizientenvektoren H1, H2, H3, welche zur Bildung des fortlaufend adoptierten Ausgangssignals y-R benötigt wurden, existiert nun ein zusätzliches Ausgangssignal (y-Ro), das unter Verwendung gespeicherter Koeffizienten H10, H20, H30 gebildet ist. Die aktueller Koeffizientensätze H1, H2, H3 stellen nur bei fehlender Spracheingabe im eingeschwungenen Zustand ein brauchbares Kompressionsfilter im Frequenzbereich dar, liefern hingegen bei Spracheingabe ungünstigende Filtereigenschaften, weil der Adaptionsprozeß in der Regelsschleife ständig gestört wird. Bei fehlender Spracheingabe d.h. hoher Filterqualität sind die drei Schalter geschlossen und es werden die aktueller Koeffizientensätze in die Koeffizientenspeicher M1, M2, M3 geschrieben: H10=H1, H20=H2, H30=H3. Die Ausgänge (y-Ro) und (y-Ra) sind identisch. Einsetzerde Spracheingaben bewirken ein Öffnen der 3 Schalter, wodurch die zuletzt in den Speichern M1, M2, M3 befindlichen Koeffizienten H10, H20, H30 nicht mehr überschrieben werden und unverändert bleiben. Diese

5 **[0032]** Als Sprachpausenmerkmal *fea* hat sich die
Ra) unterschieden, wird solange beibehalten, bis wie-
der eine Sprachpause detektiert und die Schalter
geschlossen werden.

10 **[0032]** Als Sprachpausenmerkmal *fea* hat sich die
geglättete Summe aller Absolutwerte der Koeffizienten-
koeffizientenvektoren $\Delta H1'$, $\Delta H2'$, $\Delta H3'$ bewahrt (Fig. 8a).
Diese Größe ist gleich Null bzw. weist kleine Zahlen-
werte auf, wenn es keinen oder nur einen geringen
Bedarf gibt, die Koeffizienten abzuändern. In Sprach-
pausen ist dies der Fall, der Regelkreis ist praktisch ein-
geschwungen. Störungen, wie sie durch
Spracheingabe - aber auch durch Bewegungen der
Fahrzeuginsassen - hervorgerufen werden, haben
einen erhöhten Nachregelbedarf zur Folge, was sich
durch entsprechend große Zahlenwerte bei $\Delta H1'$, $\Delta H2'$,
 $\Delta H3'$ und somit beim Merkmal *fea* bemerkbar macht.
Ein GlättungsfILTER beispielweise ein rekursiver Tieflpaß

15 1. Ordnung mit dem Eingang *fea* stellt an seinem Aus-
gang das geglättete Sprachpausen-Merkmal *fea* zur
Verfügung, welches nach Vergleich mit einem Schwell-
wert *th* die Schalter für die Koeffizientenübernahme
steuert.

20 **[0033]** Die Wirkungsweise des verbesserten FLMS-
Algorithmus nach Fig. 8 demonstriert Fig. 13. Oben ist
das aufgezeichnete Signal *y* der Szene Z1 (vgl. Fig. 11
oben) dargestellt, unten das gewonnene Ausgangssig-
nal. Schon der visuelle Vergleich der Ausgangssignale
von Fig. 13 und Fig. 11 zeigt die verbesserte Herausar-
beitung der Sprachpassagen. Der vergleichende Hör-
test bestätigt dies: auch während der Spracheingabe ist
die Musikunterdrückung deutlich besser. Den Verlauf
des Sprachpausenmerkmals und der konstanten
Schwelle über der Zeit (hier in FFT-Blöcken skaliert)
zeigt Fig. 14 oben. In den durch die Schwellenunter-
schreitung detektierten Sprachpausen (Fig. 14 unten)
findet lautend die Übernahme der Koeffizienten in die
Speicher wie beschrieben statt, um dort während der
Spracheingabe als gespeicherte Koeffizienten zur Ver-
fügung zu stehen. Die schon in Fig. 12 am Szenenende
gemessene 384-sample-Impulsantwort mit zugehöriger
Betragstransfunktion ist in Fig. 15 als aktuelle
Impulsantwort (a) bzw. aktuelle Übertragungsfunktion
(b) dargestellt. Im Gegensatz zu dieser intakte Sprach-
eingabe stark gestörten Schätzung aus den aktuellen
Koeffizienten $H1$, $H2$, $H3$ ist aus den gespeicherten
eine Übertragungsfunktion (d) hoher Qualität berechen-
bar. Die Impulsantwort aus den gespeicherten Koeffizi-
enten weist die typischen Nullsamples am Anfang auf,
welche durch die Laufzeit des Direktschalls vom Radio-
lautsprecher zum Spracheingabemikrofon verursacht
werden. Aus der im Beispielfall abzulesenden Totzeit von ca. 40 samples läßt sich die Entfernung zwischen
Lautsprecher und Mikrofon bestimmen.

25 **[0034]** Wie vorstehend schon angedeutet läßt sich die
aufwendige Projektion P4 (FFT, Fenster rechts im Zeit-
bereich, FFT) ohne merkliche Einbuße an Qualität

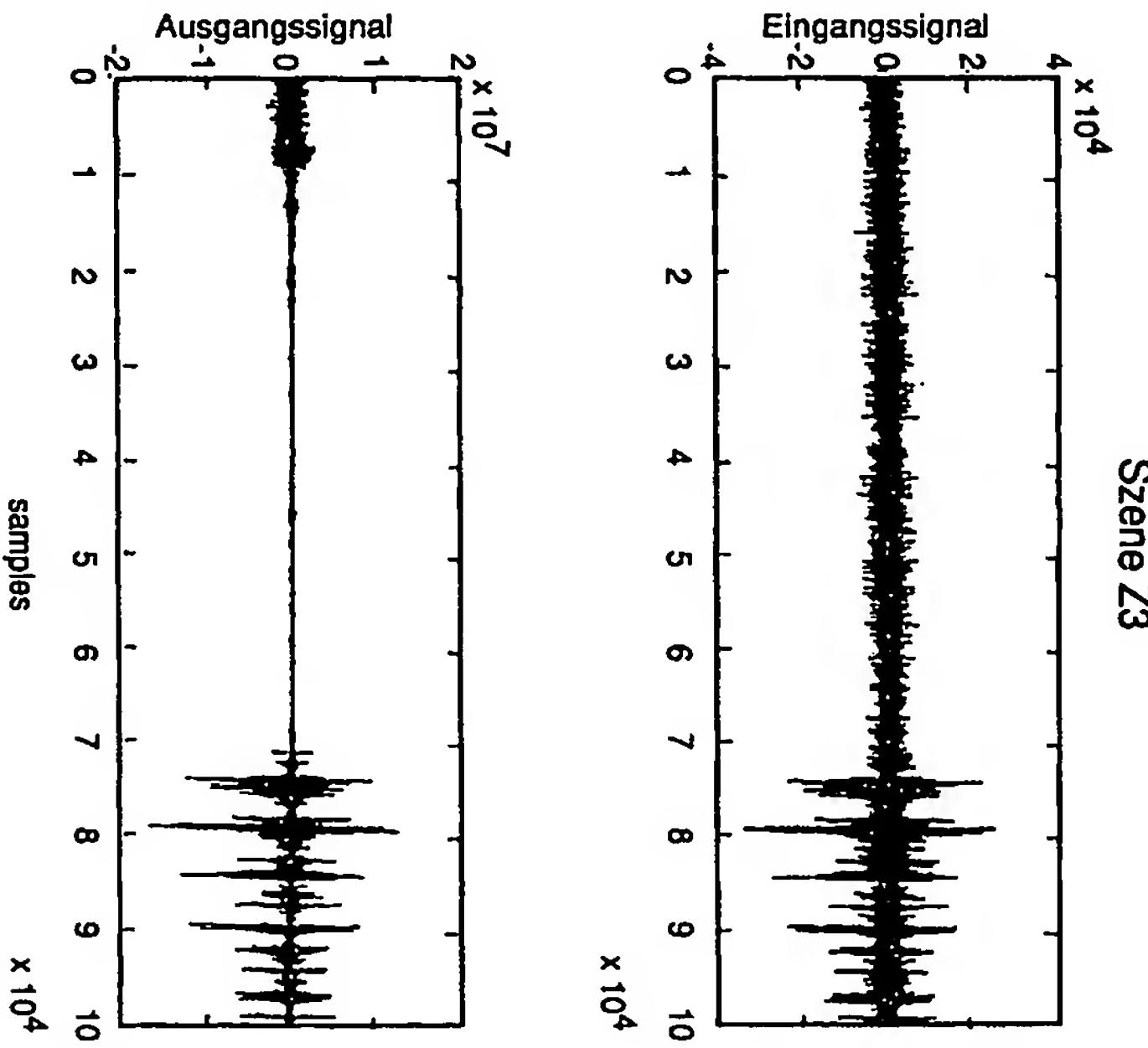
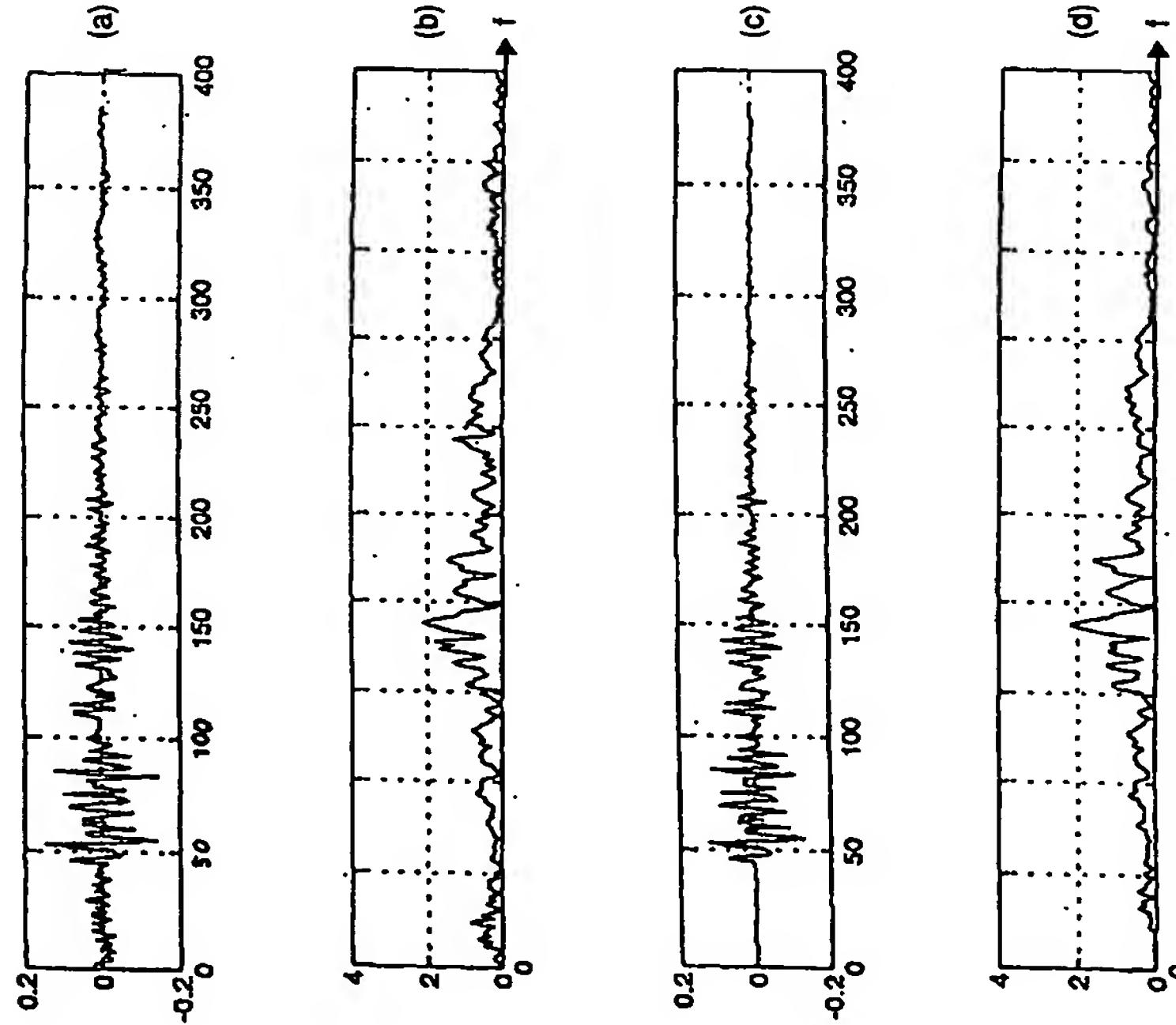


FIG. 23

durch eine relativ einfache Faltung im Frequenzbereich ersetzen, wodurch 2 FFTs eingespart werden. Dazu betrachte man Fig. 16. In einem ersten Schritt wird das "rechtsseitige" 128-sample-Rechteckfenster im Zeitbereich (Fig. 16a) bei der idealen Projektion ersetzt durch ein 128-sample-Hammingfenster (Fig. 16b). Gegenüber dem Rechteckfenster besitzt dieses den Vorteil eines bedeutend schmaleren Spektrums. Wie Fig. 17 zeigt, besteht beim Rechteckfenster der Realteil des Spektrums aus einer einzigen Linie (Gleichanteil), während das zur Mitte antisymmetrische Imaginarteil-Spektrum aus vielen nach außen hin langsam abfallenden Linien mit alternierenden Nullstellen besteht. Im Gegensatz dazu beschränkt sich das komplexe Spektrum des Hammingfensters (Fig. 18) auf insgesamt 7 Linien, von denen im symmetrischen Realteil nur 3 und im antisymmetrischen Imaginarteil nur 4 Werte von Null verschiedenen sind. Sämtliche weiter außen liegenden Anteile sind vernachlässigbar gering. Diese spezielle Eigenschaft des Hammingfensters ermöglicht es vorteilhaft, teilweise die Multiplikation im Zeitbereich (Fig. 16b) zu ersetzen durch eine Faltung mit dem zugehörigen 7-

Szene Z2



22

Faltung im Frequenzbereich (Fig. 18c) mit einem komplexen 7-Linien-Spektrum (Fig. 18) sei anhand von 3 Meßzahlen demonstriert.

[0037] Die erste dieser Szenen 22 beinhaltet Spracheingabe von Ziffern, wobei der Radiolautsprecher amhend weißes Rauschen mit verhältnismäßig hoher Lautstärke abstrahlt. Das zugehörige 100000-sample-Mikrophonsignal ist in Fig. 20 oben, das extrahierte Ausgangssignal in Fig. 20 unten dargestellt. Eine deutliche Rauschentfernung des Outputsignals gegenüber dem Mikrofoninput stellt man durch Abhörvergleich fest. Der zeitliche Verlauf des Sprachpausenmerkmals ist zusammen mit der konstanten Schwellw. in Fig. 21 oben abgebildet und die hieraus abgeleiteten Sprachpausen bzw. die zugeordneten Schalterstellungen in Fig. 21 unten. Schließlich zeigt Fig. 22 in zu Fig. 15 analoger Weise die am Szenenende gefundene Impulsantwort (a) und Übertragungsfunktion (b) auf der Basis der aktuellen Koeffizienten und die entsprechenden Größen (c), (d) auf der Basis der Sprachpauseneinstellung. Es ist deutlich erkennbar, daß die am Szenenende gefundene aktuelle Impulsantwort ein infolge Spracheingabe gestörtes Ergebnis darstellt, während die aus der letzten Sprachpause stammende Impulsantwort aus den errechneten Koeffizientenwerten eine hohe

[0038] Die ersten 100000 samples einer Messszene Z3 mit POP-Musik im Radio und flüssig bis schnell gesprochener Sprache der rechts hinten sitzenden Personen sind in Form des Mikrophonsignales y in Fig. 23 oben aufgezeichnet. Nach ca. 10000 samples (0.83 s) wird das Radiosignal brauchbar unterdrückt (Fig. 23 unten). Auch bei der im letzten Drittel dieser Szene einsetzenden Spracheingabe bleibt die POP-Musikunterdrückung wirksam erhalten, wodurch die Sprachverständlichkeit hier gegenüber dem Mikrophonsignal merklich verbessert wird. Nach einer langen Sprachpause kommt es wegen der anschließenden pausenfreien Spracheingabe nicht mehr zu einer Schwellenumstreuung (Fig. 24). Aus diesem Grunde ist die in Fig. 25 unten am Ende der Szene festgehaltene Impulsantwort auf der Basis der gespeicherten Koeffizienten zeitlich relativ veraltet, weil sie bereits ca. 2.3 s vorher aktuell war (215 Blöcke \cdot 10.7 ms). Weider weist die aktuelle Impulsantwort (Fig. 25 oben) starke von der Spracheingabe hervorrende Störungen auf. Wie ein Vergleich mit der ähnlichen Szene Z1 nach Figuren 11 bis 15 zeigt, ist trotz des stark verringerten Rechenaufwandes die Qualität der Stoßbeseitigung unverändert hoch.

[0039] Die letzte Szene Z4 nach Fig. 26 wurde ohne Spracheingabe erstellt und soll abschließend nochmals die Musikunterdrückungseigenschaften des beschriebenen FUMS-Algorithmus demonstrieren. Nach ca. 1800 samples bzw. 1.5 s wird - wie aus Fig. 26 unten ersichtlich - die Musik wirksam unterdrückt. Diese Eigenschaft wird bis zum Szeneende mit unveränderter Qualität beibehalten. Ein 27. rotes auf dem

durch eine relativ einfache Faltung im Frequenzbereich ersetzen, wodurch 2 FFT's eingespart werden. Dazu betrachte man Fig. 16. In einem ersten Schritt wird das 'rechtsseitige' 128-sample-Rechteckfenster im Zeitbereich (Fig. 16a) bei der idealen Projektion ersetzt durch ein 128-sample-Hammingfenster (Fig. 16b). Gegenüber dem Rechteckfenster besitzt dieses den Vorteil eines bedeutend schmaleren Spektrums. Wie Fig. 17 zeigt, besteht beim Rechteckfenster der Realteil des Spektrums aus einer einzigen Linie (Gleichanteil), während das zur Mitte antisymmetrische Imaginärteil-Spektrum aus vielen nach außen hin langsam abfallenden Linien mit alternierenden Nullstellen besteht. Im Gegensatz dazu beschränkt sich das komplexe Spektrum des Hammingfensters (Fig. 18) auf insgesamt 7 Linien, von denen im symmetrischen Realteil nur 3 und im antisymmetrischen Imaginärteil nur 4 Werte von Null verschiedenen sind. Sämtliche weiter außen liegenden Anteile sind vernachlässigbar gering. Diese spezielle Eigenschaft des Hammingfensters ermöglicht es vorteilhaftweise die Multiplikation im Zeitbereich (Fig. 16b) zu ersetzen durch eine Faltung mit dem zugehörigen 7-sample-Spektrum im Frequenzbereich und damit eine IFFT und eine FFT einzusparen (Fig. 16c).

tion P1 (IFFT - linksseitiges Rechteckfenster - FFT) 30 55
ersetzen durch eine entsprechende Faltungsoperation im Frequenzbereich mit dem konjugiert komplexen 7-Linier-Spektrum. Experimente haben jedoch gezeigt, daß Einsparungen an dieser Stelle erkauft werden mit einer deutlichen Verschlechterung des Einschwingverhaltens. Aufwandsünstige Lösungen lassen sich trotzdem dadurch erzielen, daß in dem LMS-Algorithmus nach Fig. 8 die 3 Projektionen P1 nicht gleichzeitig in einem 256-sample-Inputdatenblock abgearbeitet werden müssen. Die mit 128-samples überlappenden Inputdatenblöcke der Länge 256 sind mit einer willkürlich bei "1" beginnenden Nummerierung in Fig. 19a skizziert. So ist es z.B. möglich bei modulo-3-Zahlweise der Inputdatenblöcke die 3 Teilfilterprojektionen nicht parallel (Fig. 19b) sondern sequentiell in aufeinanderfolgenden Blöcken Fig. 19 zu berechnen. Dadurch sind bei idealer Projektion P1 pro Datenblock nicht 6 sondern nur noch 2 FFT's notwendig. Es hat sich gezeigt, daß die Komplexierung des Radiosignals auch noch ausreichend funktioniert, wenn die Abstände zwischen den zu berechnenden Teilfilter-Projektionen noch größer gewählt werden. Zählt man die Blöcke z.B. modulo 6, so ist lediglich in jedem zweiten Block eine Projektion zu berechnen (Fig. 19d). Selbst eine Reduzierung auf einen Abstand von vier Blöcken zwischen zwei aufeinanderfolgenden P1-Berechnungen mittels modulo-12-Zählung führt noch zu brauchbaren Ergebnissen (Fig. 19e).

[0036] Die Leistungsfähigkeit des FLMS-Algorithmus mit 3 Teilfiltern gemäß Blockschaltung Fig. 8 und einer sequentiellen Berechnung der idealen Projektion P1 im 30 35 40 45 50 55 75

Sprachpausen-Große t_{ea} überwiegend unter der Schwellen t_h bleibt. Die Zeiten, in welchen auf die gespeicherten Koeffizienten zurückgegriffen wird, sind demnach nur sehr kurz. Impulsantwort und Übertragungsfunktion aus aktuellen Koeffizienten sind daher im wesentlichen mit den entsprechenden Verläufen aus Sprachpausen-Koeffizienten identisch.

Patentansprüche

1. Verfahren zur Störbefeilung eines Mikrofoni-
grats von Anteilen eines Quellsignals, daß als Referenzsignal (X) vorliegt und nach Durchlaufen einer Übertragungsstrecke mit a priori unbekannter Übertragungsfunktion (G) sich im Mikrofoni-
grat als Störignal (I) einem Sprachsignal (S) überla-
gert, durch adaptive Nachbildung des Störsignals
und Kompenstation des tatsächlichen und des
nachgebildeten Störsignals in einem Ausgangssi-
gnal, wobei das Mikrofoni-
grat im Frequenzbereich transformiert, die Kompen-
sation im Frequenzbereich transformiert
und das im Frequenzbereich vorliegende Aus-
gangssignal zur Adaption der Nachbildung mit dem
im Frequenzbereich vorliegenden Referenzsignal
verknüpft wird.
2. Verfahren nach Anspruch 1, dadurch gekenn-
zeichnet, daß das Ausgangssignalspektrum in den
Zeitbereich transformiert, das Zeitsignal durch Vor-
anstellen von Nullwerten auf doppelte Länge
gebracht, in den Frequenzbereich rücktransformiert
und der Nachbildung der Übertragungsfunktion
zugrundegelegt wird.
3. Verfahren nach Anspruch 1, dadurch gekenn-
zeichnet, daß das Ausgangssignalspektrum mit
dem Spektrum eines Hamming-Zeitfensters geta-
tet wird und der Nachbildung der Übertragungs-
funktion zugrundegelegt wird.
4. Verfahren nach einem der Ansprüche 1 bis 3,
dadurch gekennzeichnet, daß zur Nachbildung des
Störsignalsanteils eine adaptive Filterfunktion eines
Nachbildungsfilters auf das Referenzsignal ange-
wandt wird.
5. Verfahren nach Anspruch 4, dadurch gekenn-
zeichnet, daß die Filterfunktion durch einen Koeffi-
zientenvektor vorgegeben wird, dessen
Koeffizienten adaptiv eingestellt werden.
6. Verfahren nach Anspruch 4 oder 5, dadurch
gekennzeichnet, daß das Auftreten eines Sprachsi-
gnals im Mikrofoni-
grat im Mikrofoni-
grat detektiert wird und
bei Auftreten eines Sprachsignals die vor Auftreten
des Sprachsignals eingestellte Filterfunktion zur
Bildung des Ausgangssignals beibehalten wird.

7. Verfahren nach Anspruch 6, dadurch gekenn-
zeichnet, daß auch bei Detektion eines Sprachsi-
gnals die adaptive Nachbildung einer aktuellen
Filterfunktion zusätzlich zur Bildung des Ausgangs-
signals fortgeführt wird.

8. Verfahren nach Anspruch 7, dadurch gekenn-
zeichnet, daß das Auftreten eines Sprachsignals
aus einer Veränderung der aktuellen Filterfunktion
detektiert wird.

9. Verfahren nach Anspruch 8, dadurch gekenn-
zeichnet, daß die Veränderung der aktuellen Filter-
funktion für die Detektion des Auftretens eines
Sprachsignals zeitlich geglättet wird.

10. Verfahren nach einem der Ansprüche 4 bis 9,
dadurch gekennzeichnet, daß die Filterfunktion in
mehrere Teillfilterfunktion zu aufeinanderfolgenden
Abschnitten einer Gesamt-Impulsantwort aller Teil-
filter aufgespalten ist und auf Referenzsignalspek-
trum zu zeitlich versetzten Zeitsegmenten des
segmentierten Referenz-Zeitsignals angewandt
wird.

11. Verfahren nach Anspruch 10, dadurch gekenn-
zeichnet, daß die Adaption der Filterfunktion für die
Teillfilter parallel durchgeführt wird.

12. Verfahren nach Anspruch 10, dadurch gekenn-
zeichnet, daß die Adaption der Filterfunktion für die
einzelnen Teillfilter zeitsequential durchgeführt
wird.

Szene 22

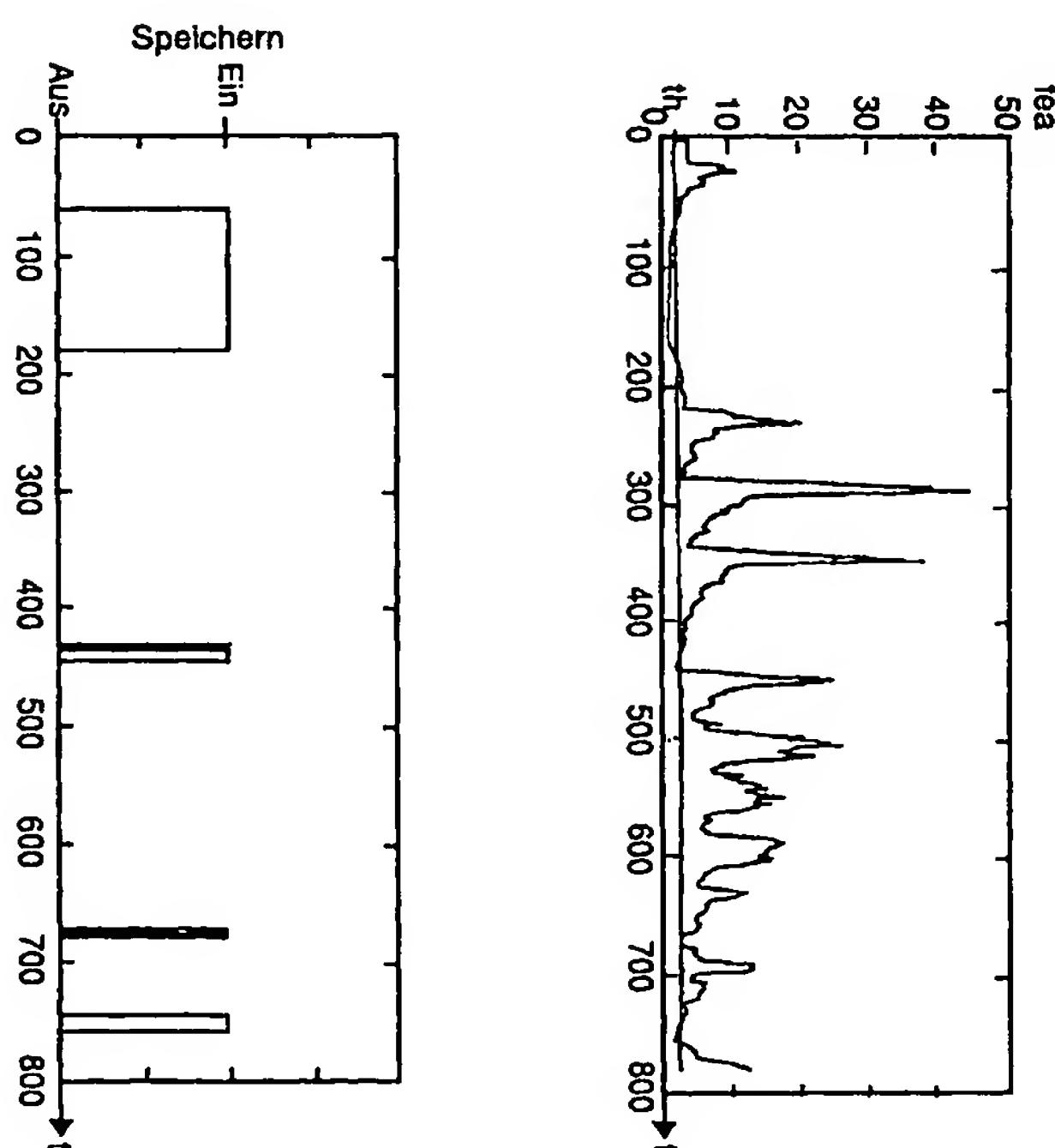


FIG.21

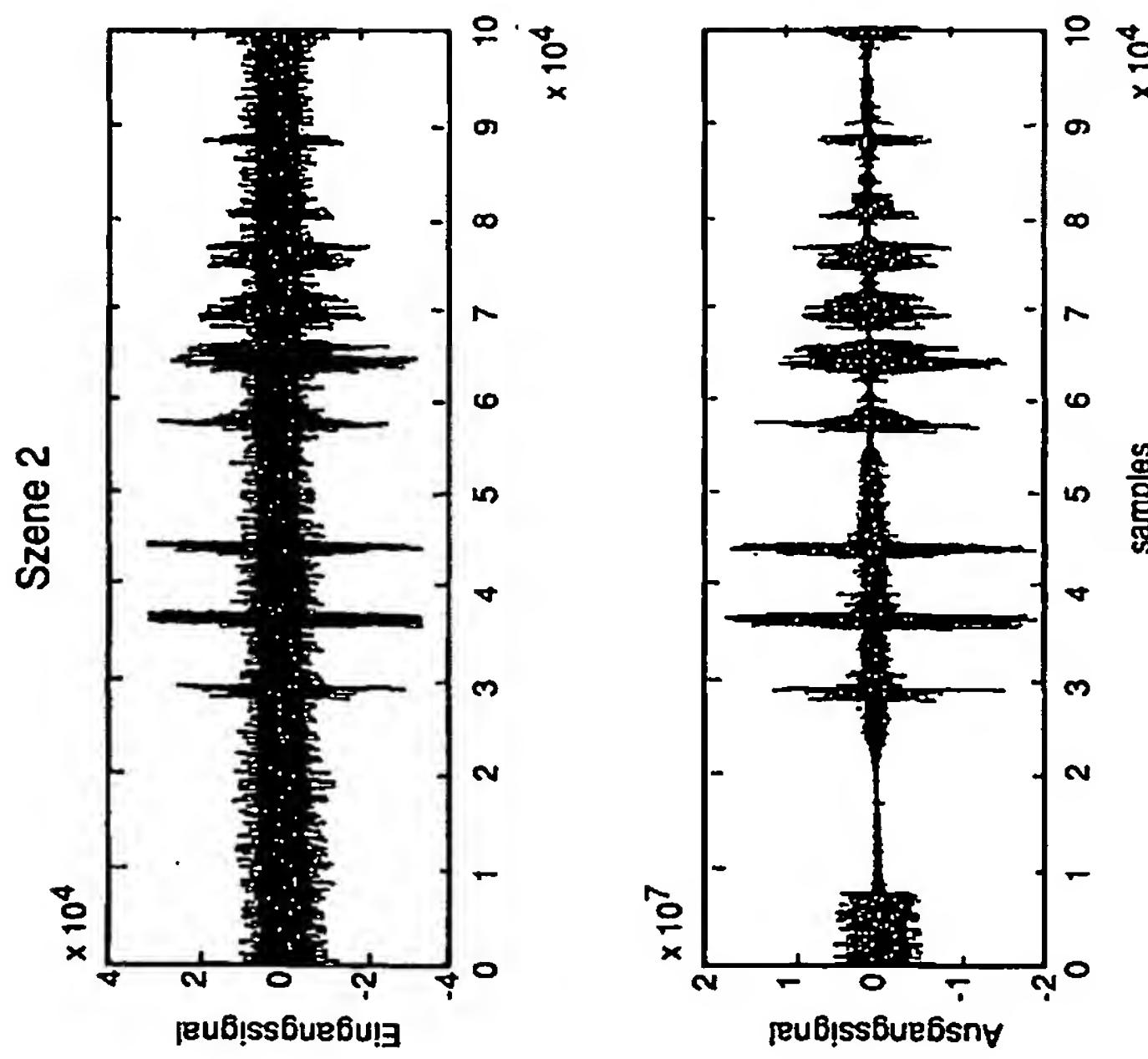


FIG.20

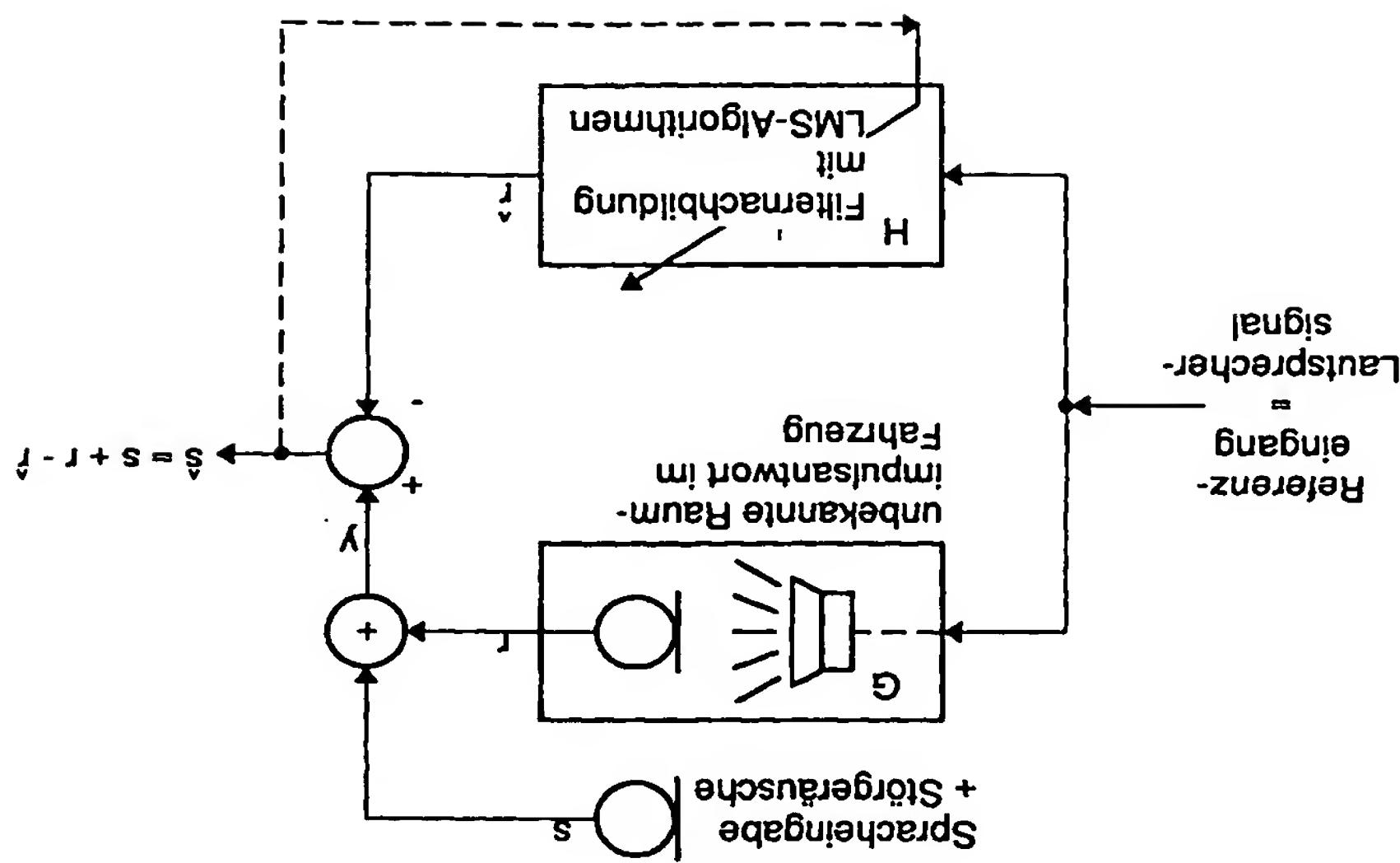
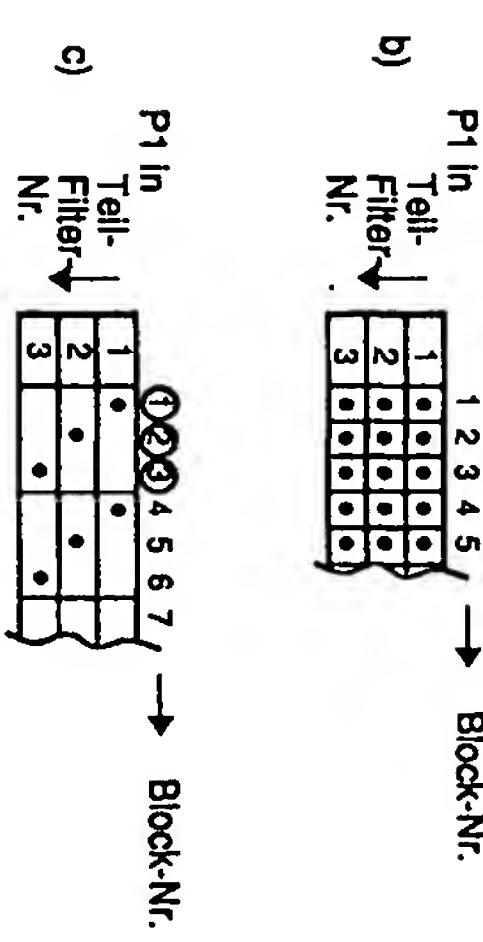
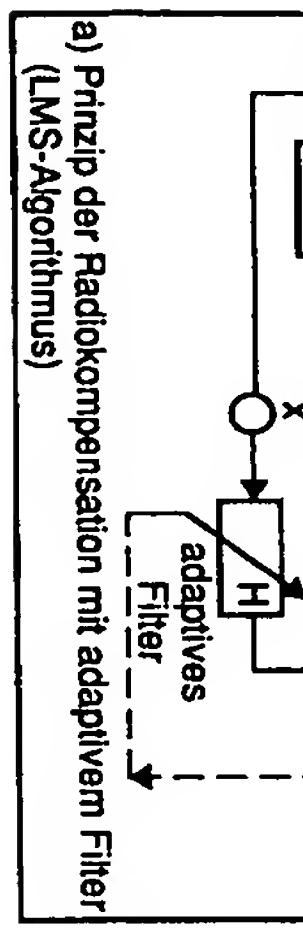
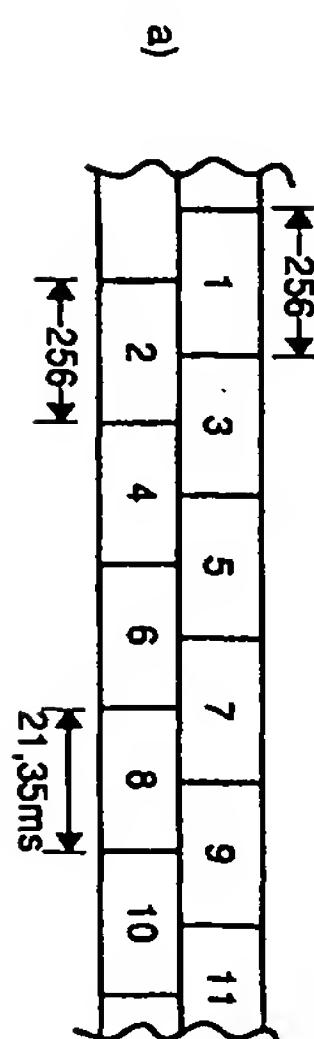


FIG.1



a) Numerierung der überlappenden FFT-Blöcke

b) Projektion P1 parallel in 3 Teilfiltern

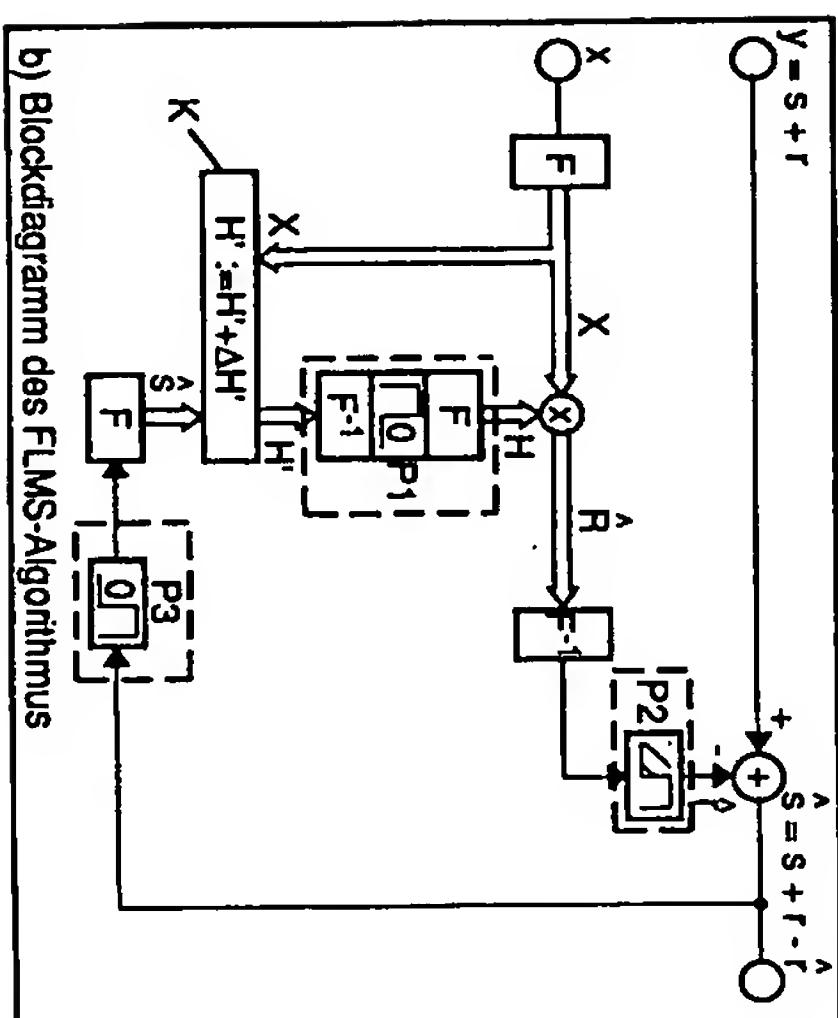
c) Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 3)

d) Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 6)

e) Projektion P1 sequentiell in 3 Teilfiltern (Blockmodule 12)

FIG.2

FIG.19



b) Blockdiagramm des FLMS-Algorithmus

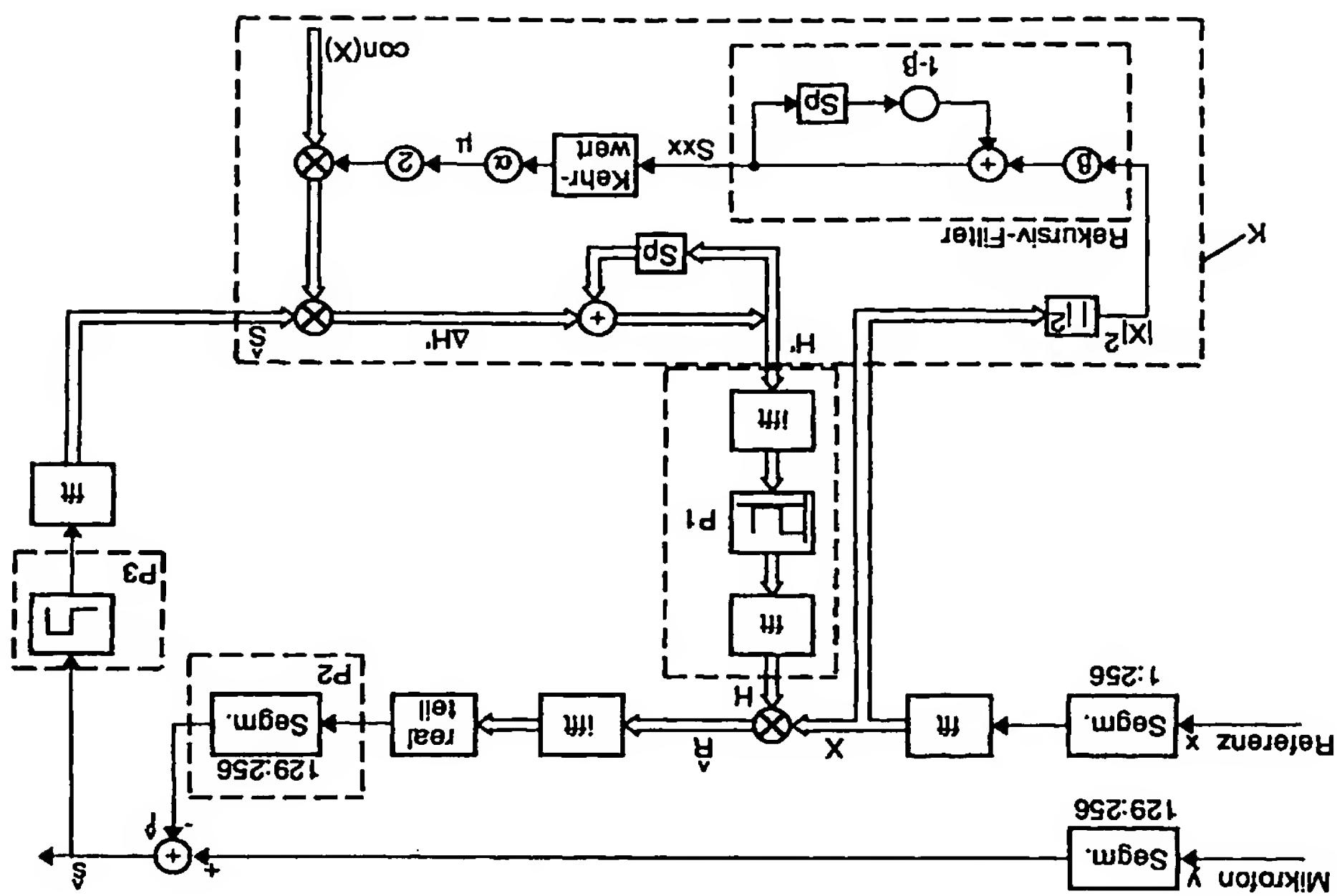


FIG.3

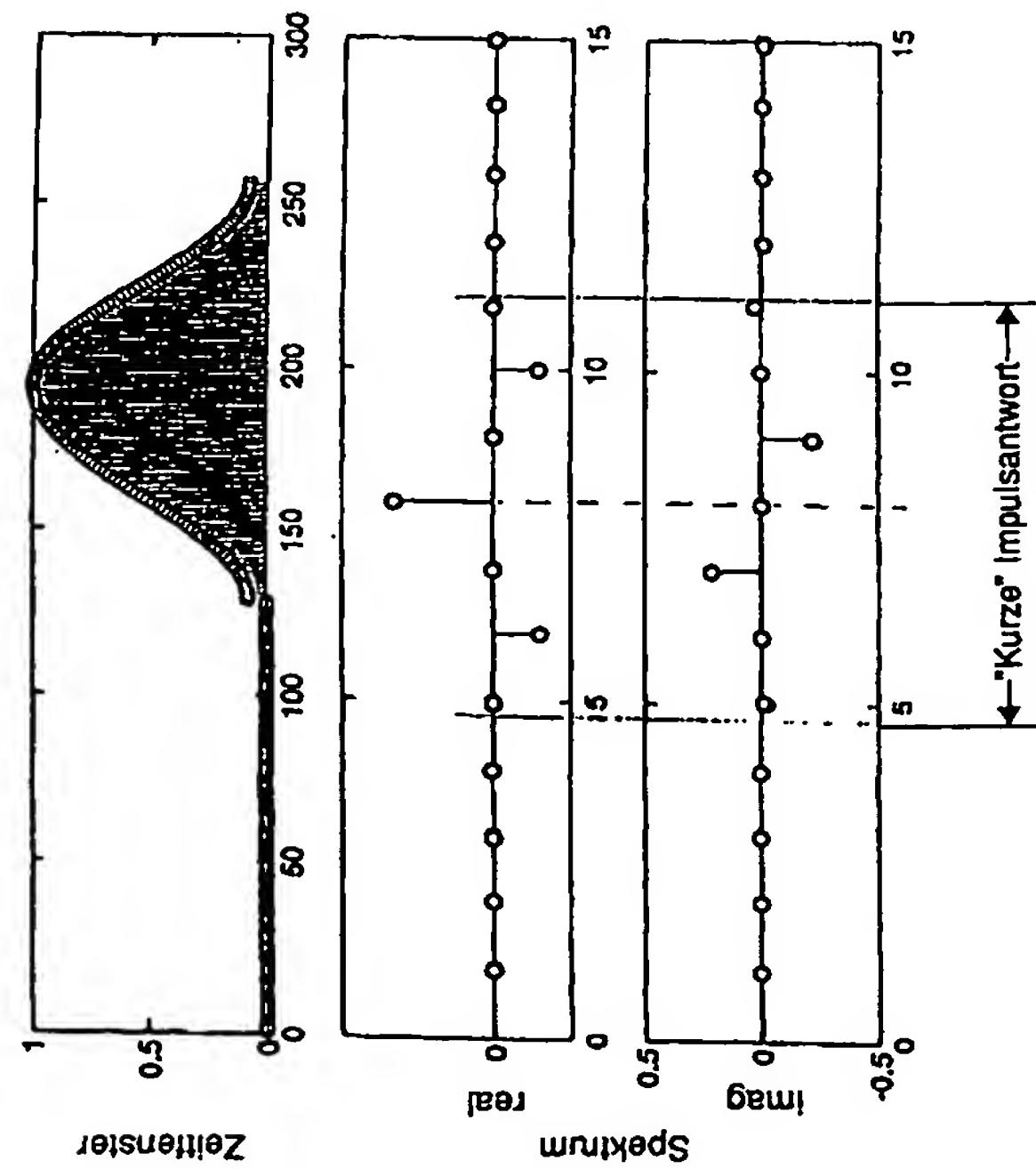
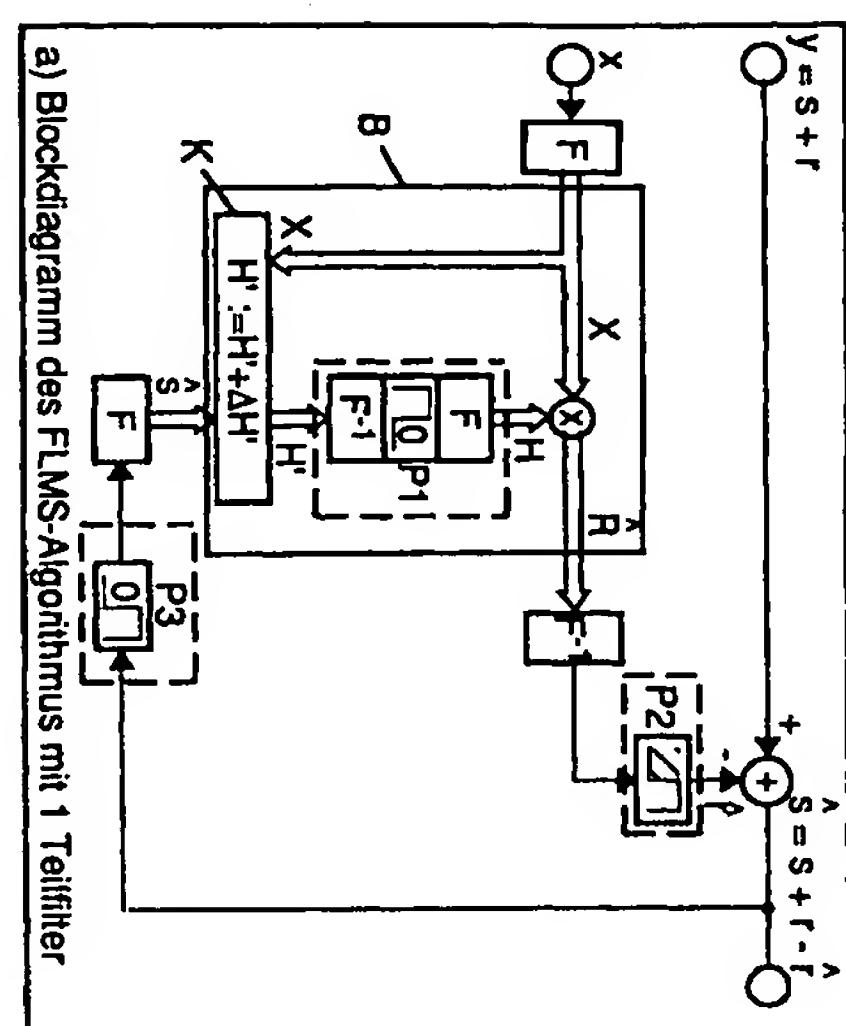
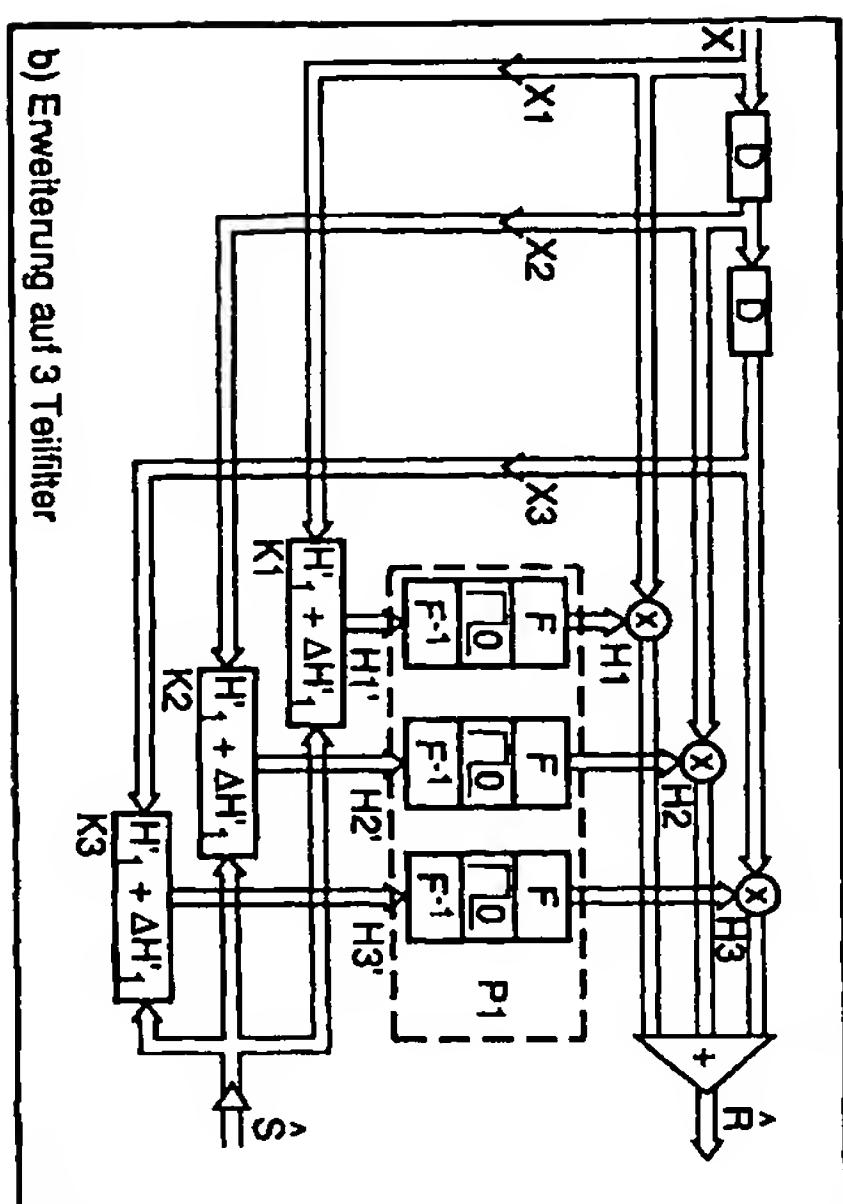


FIG.18



a) Blockdiagramm des FLMS-Algonquinus mit 1 Lamellen



b) Erweiterung auf 3 Teilträger

FIG. 4

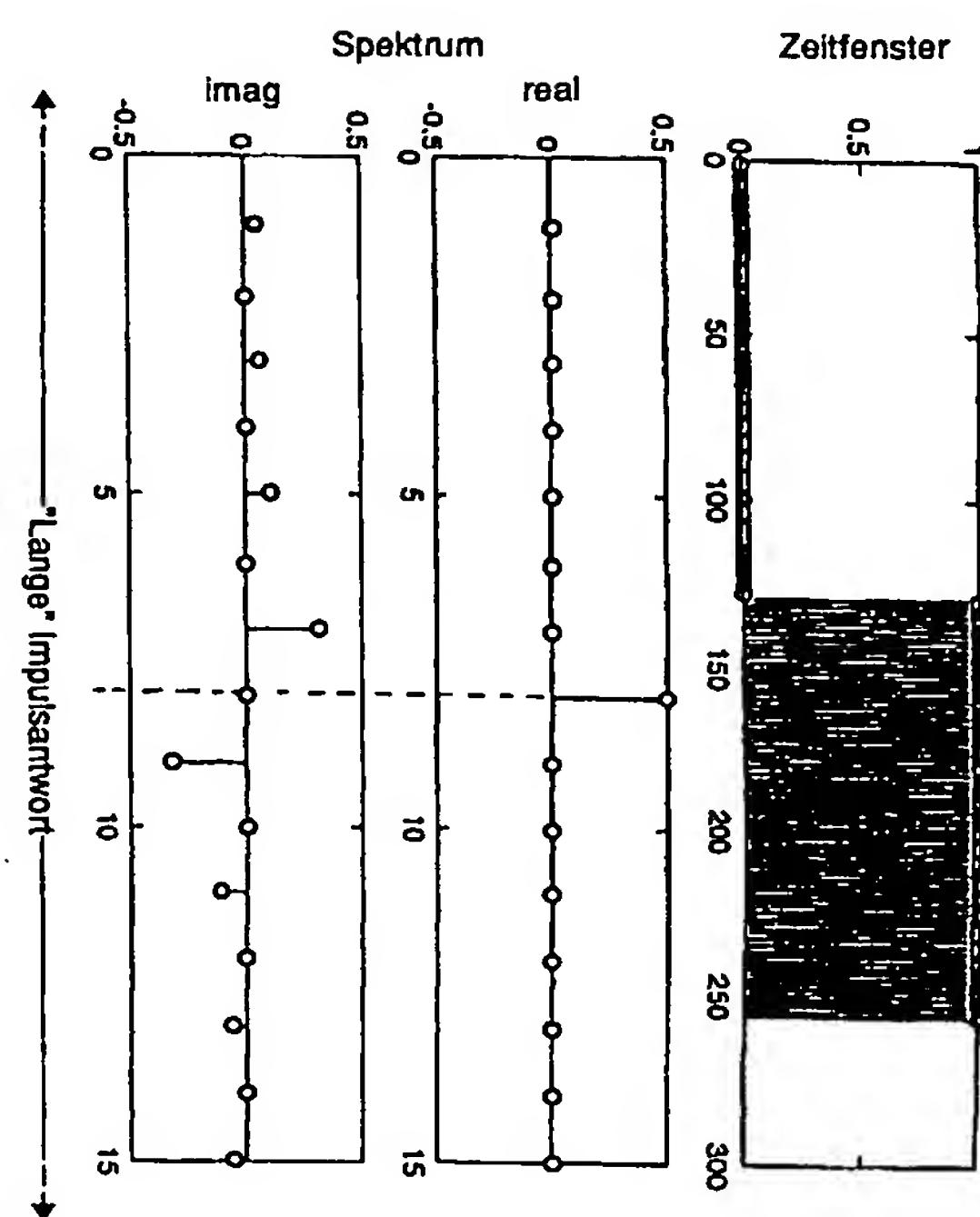
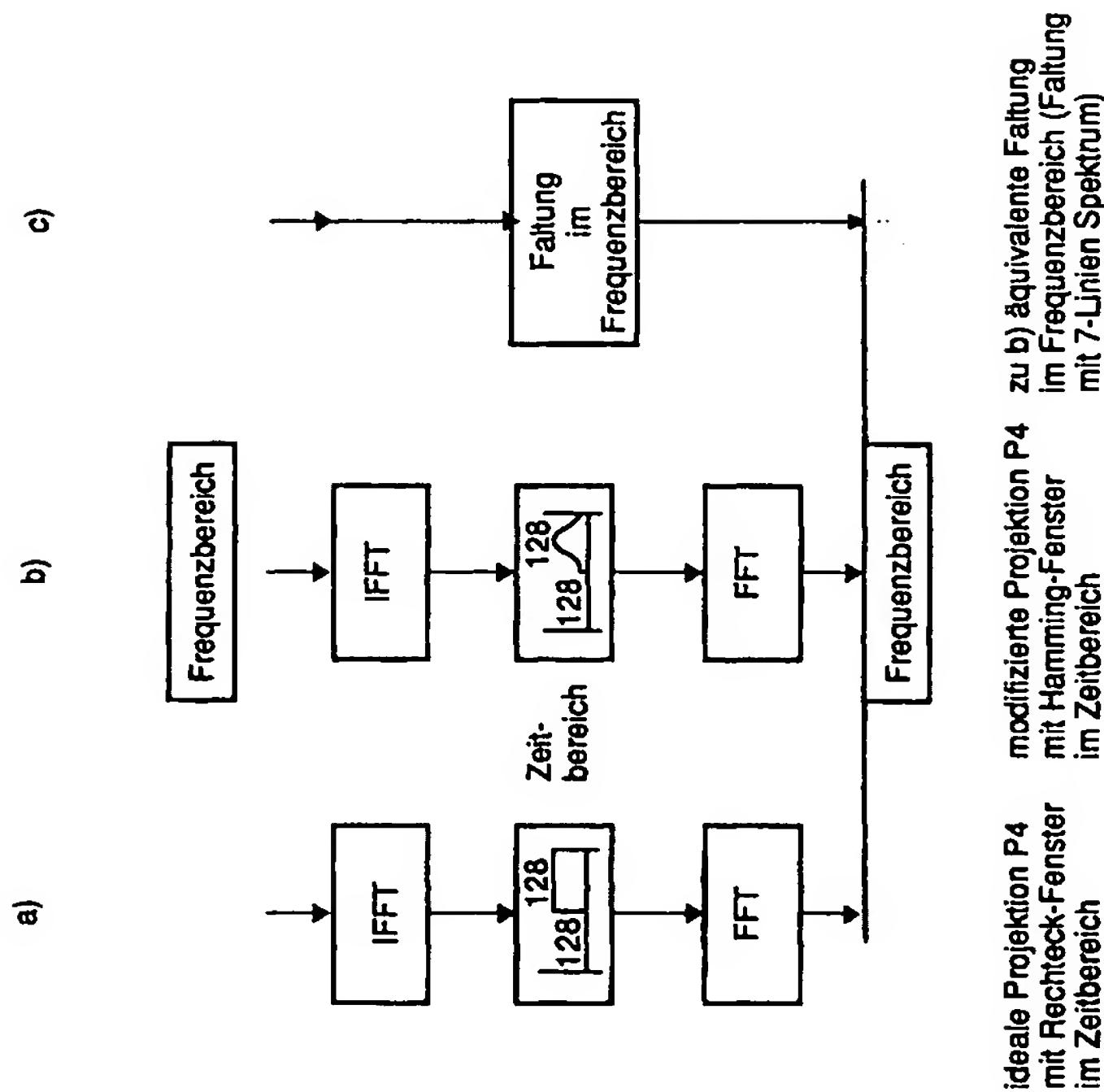
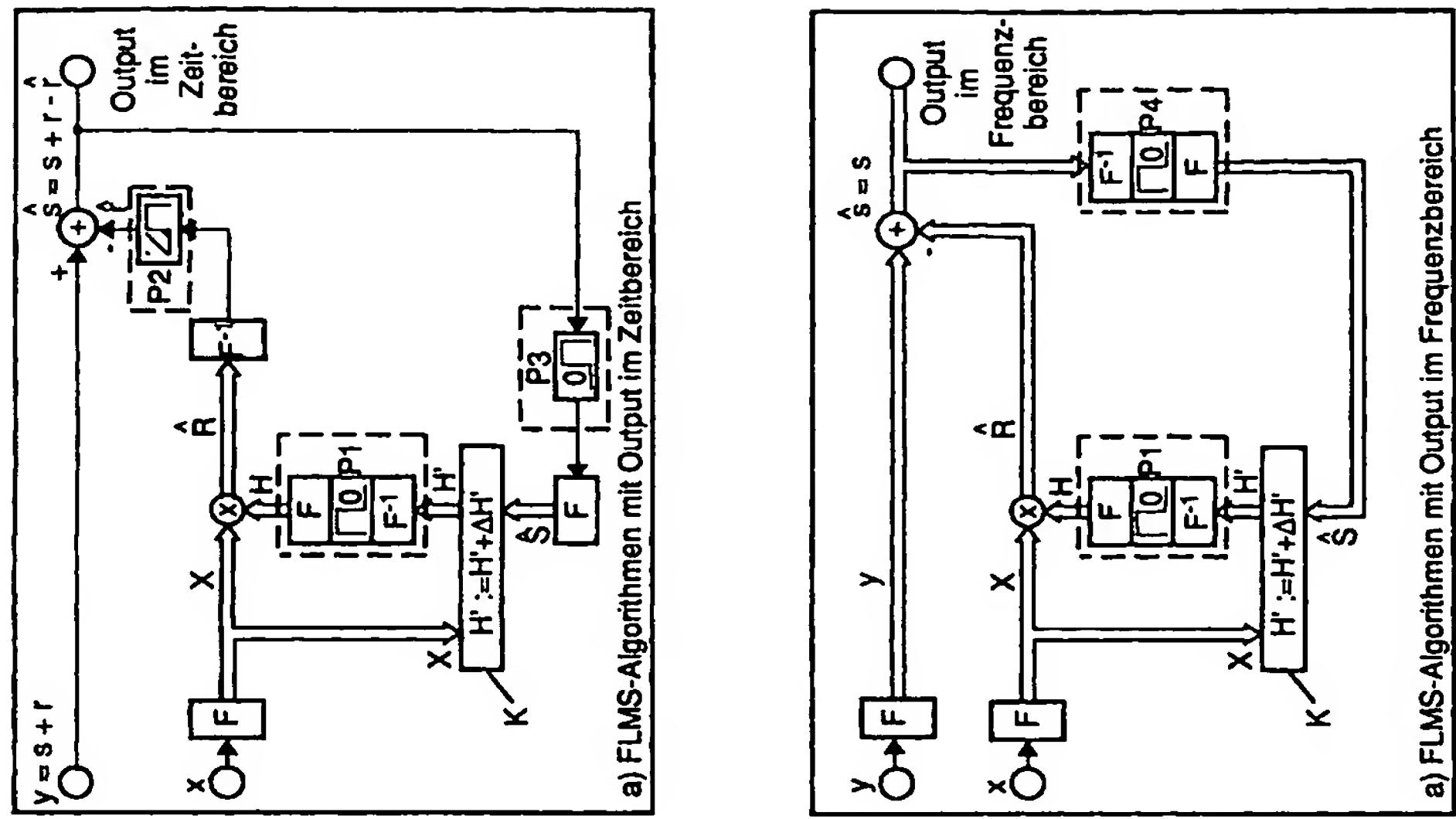


FIG. 17



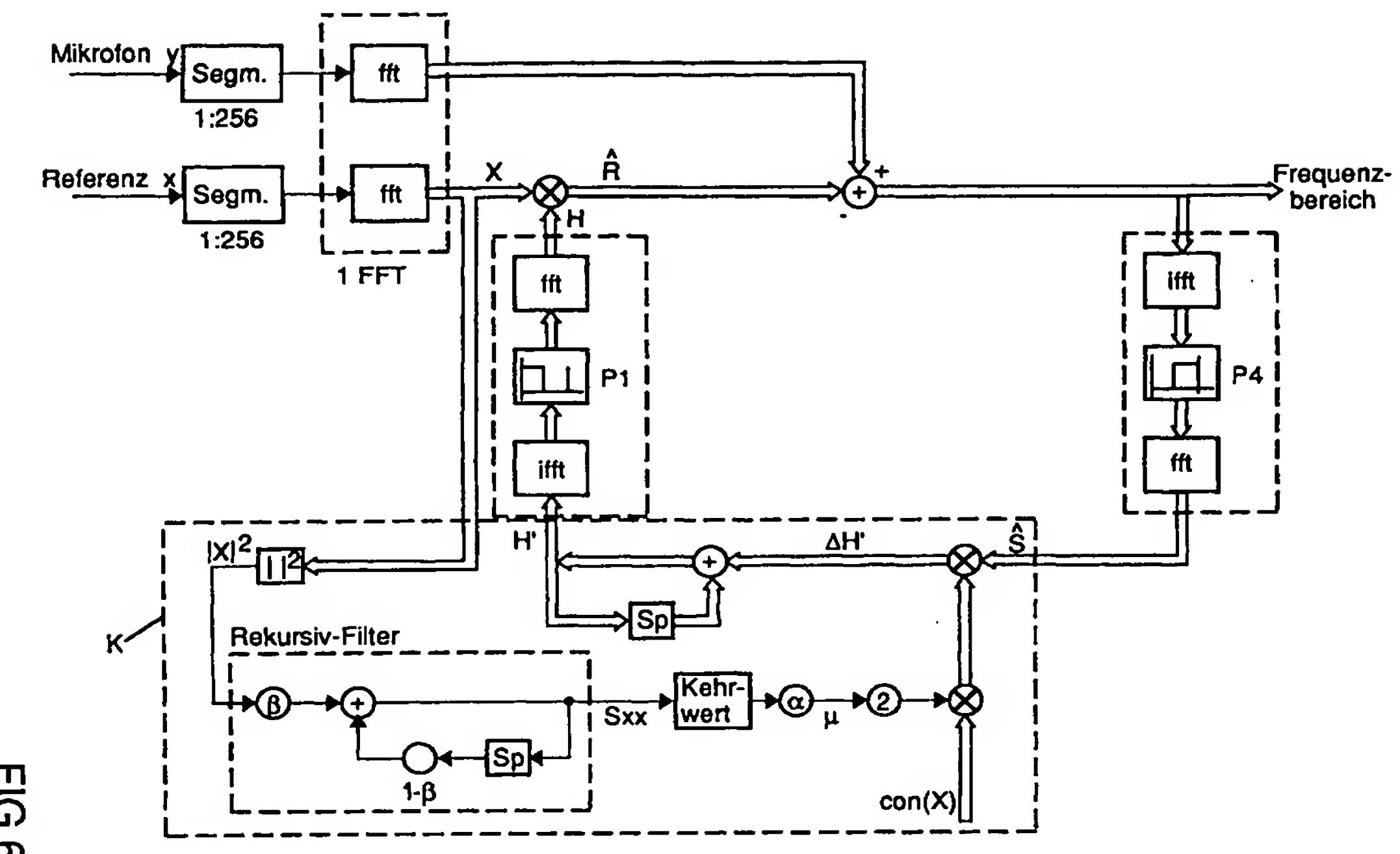


FIG. 6

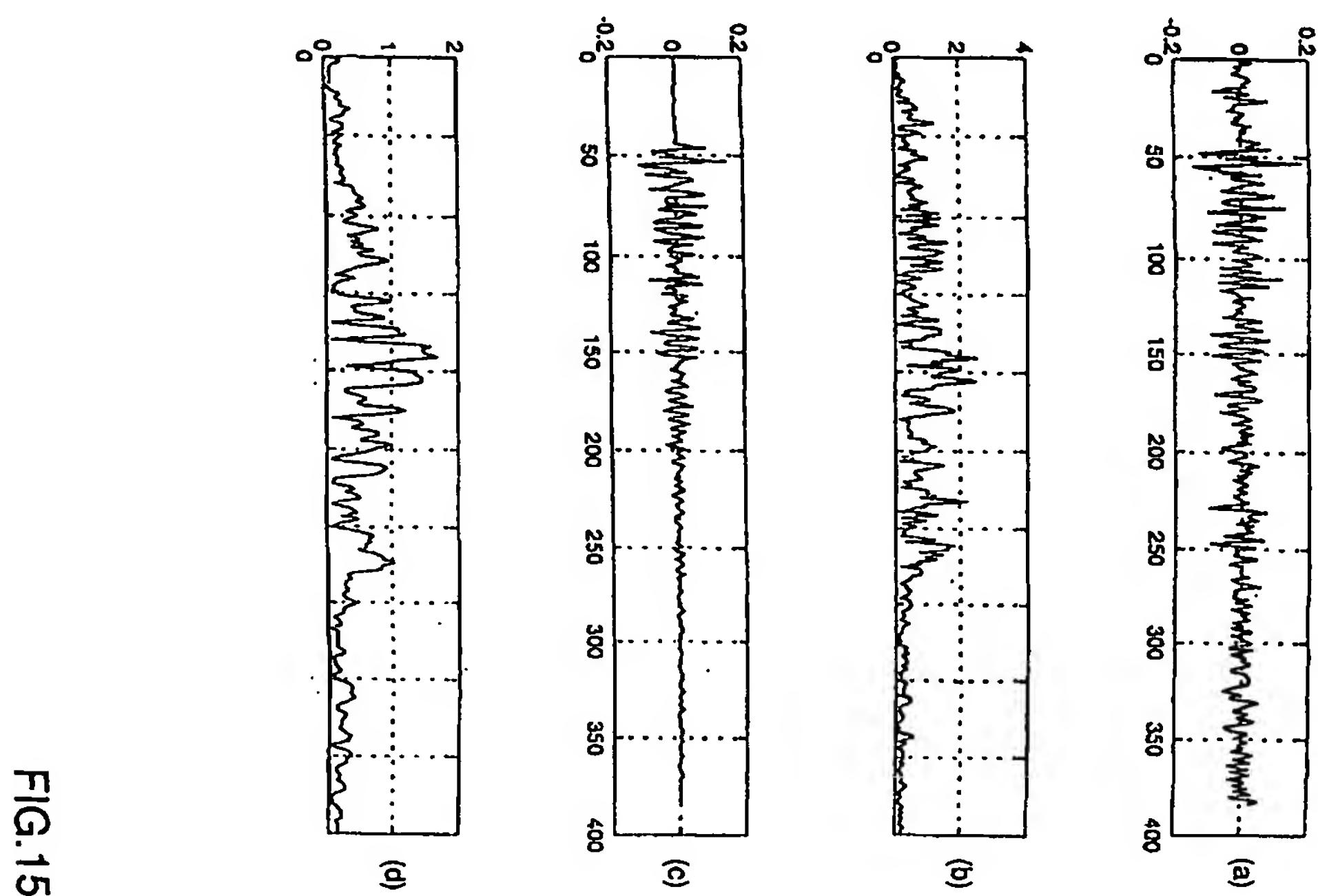


FIG. 15

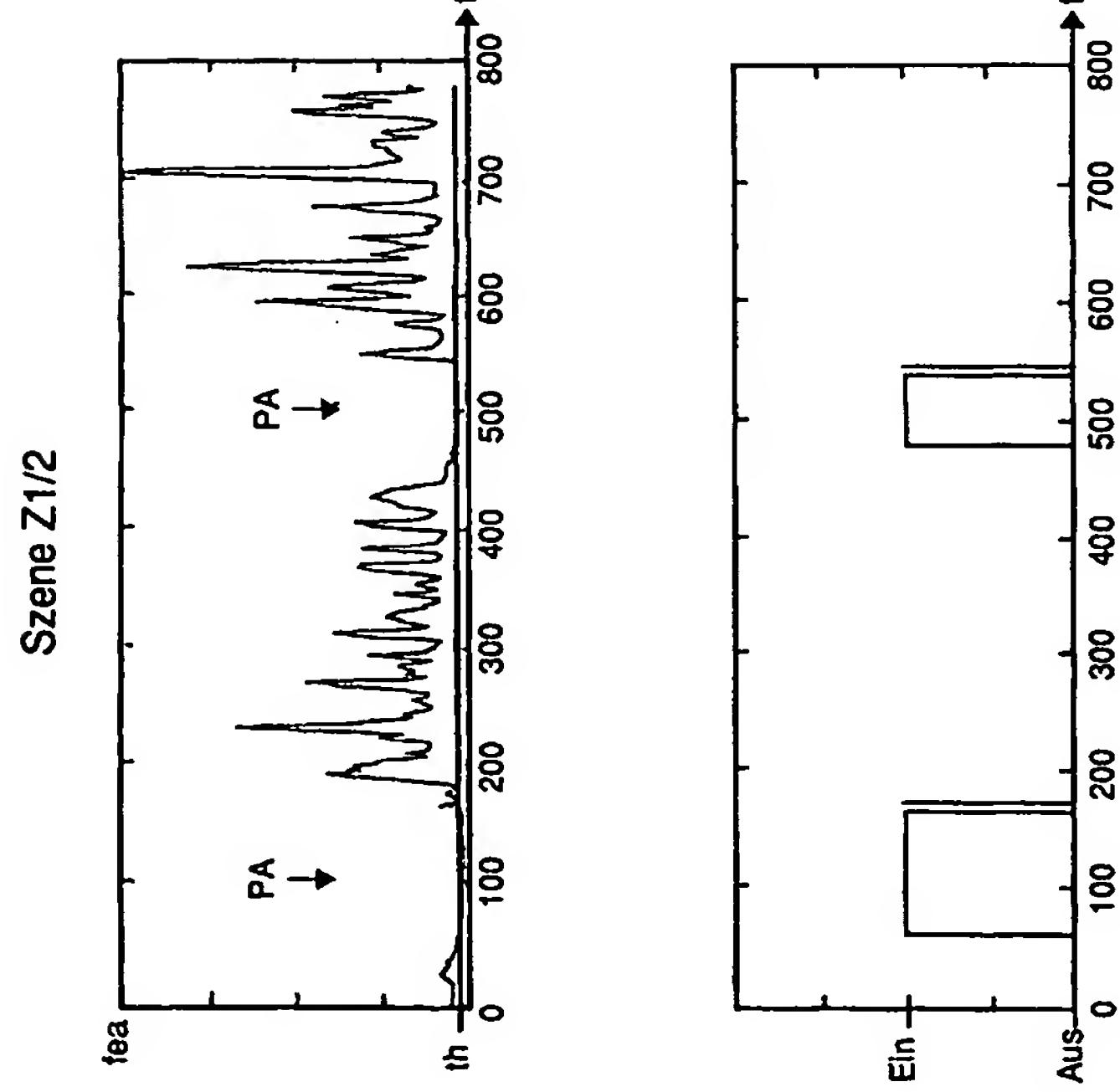


FIG.14

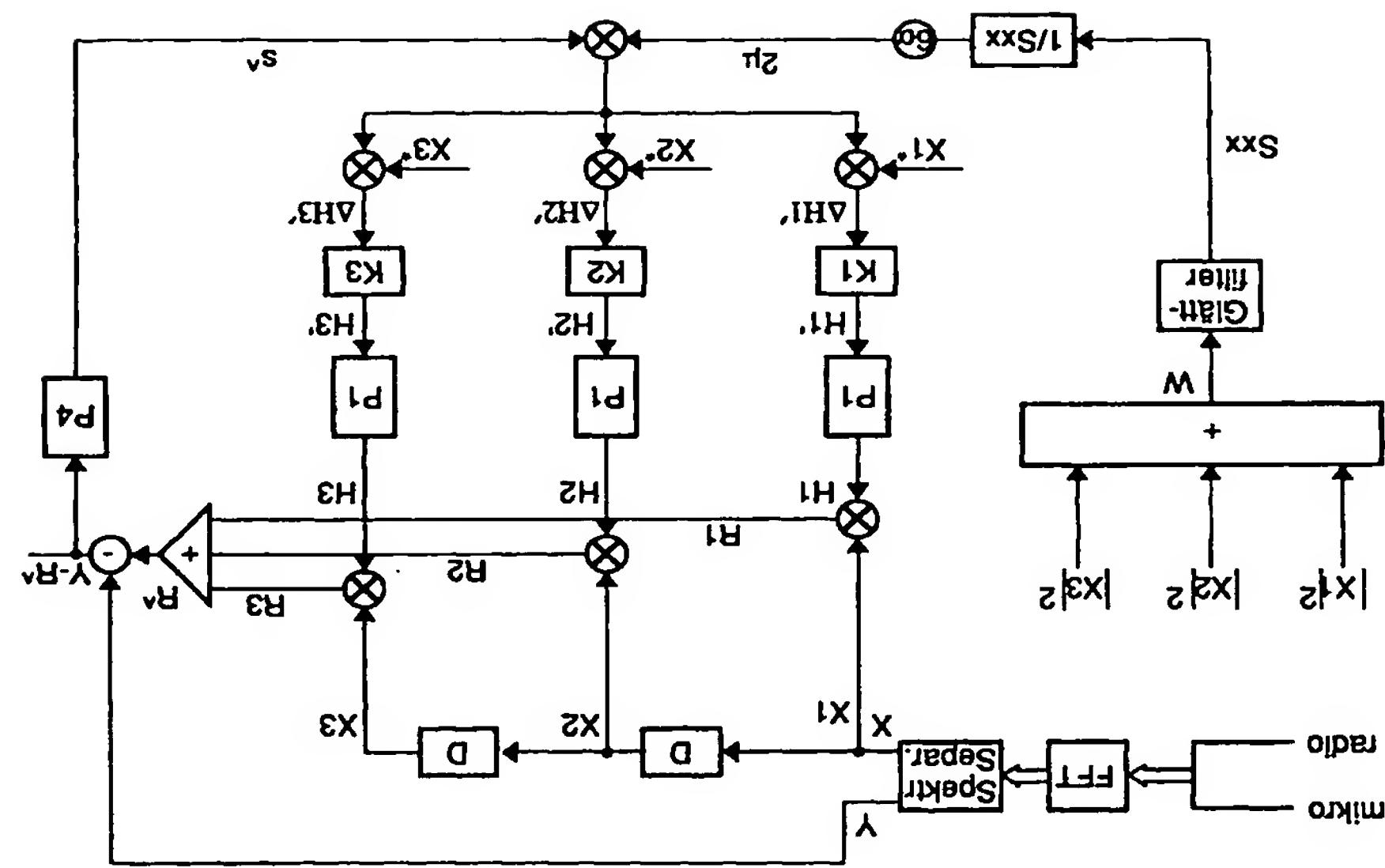


FIG.7

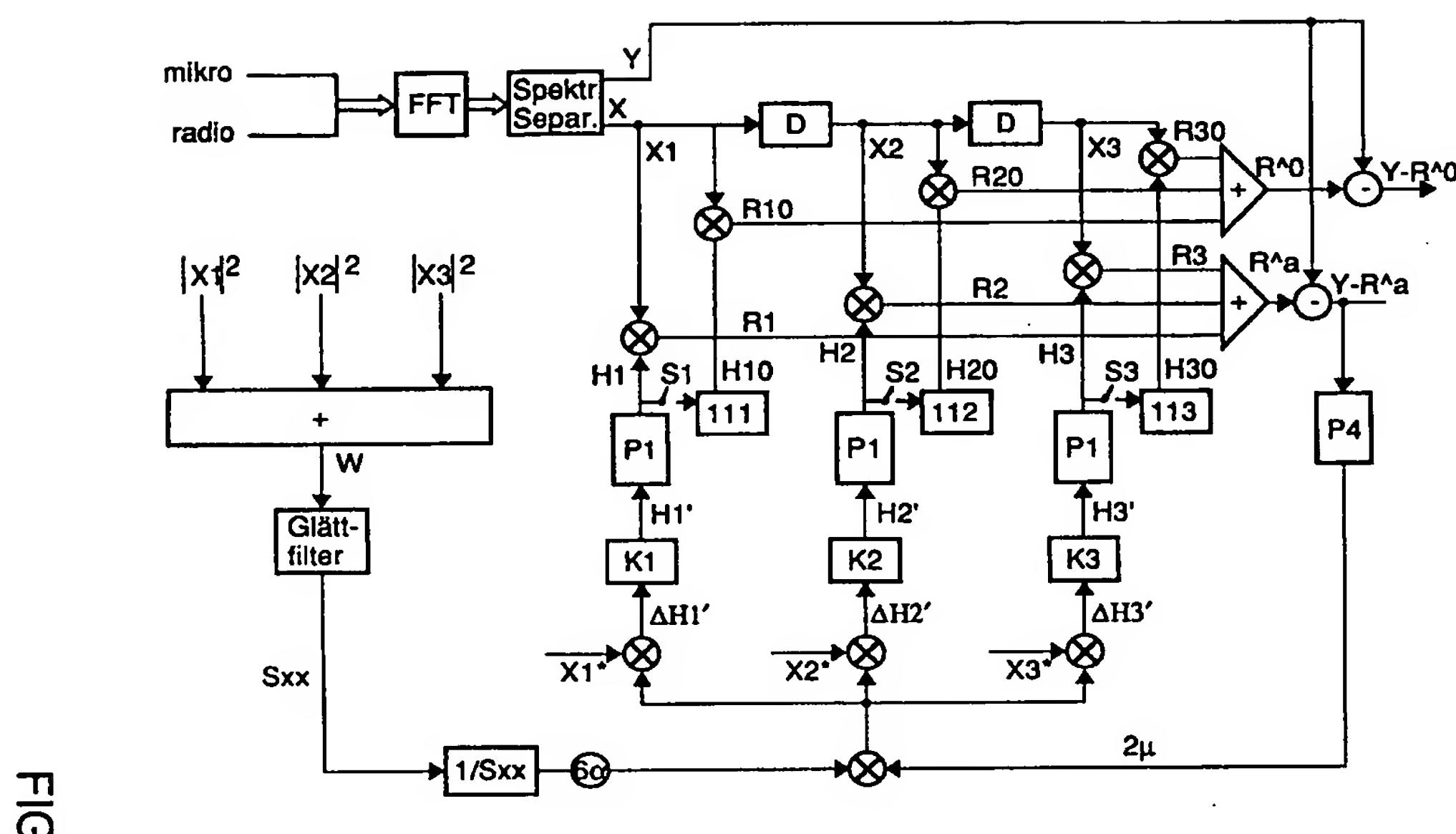


FIG.8

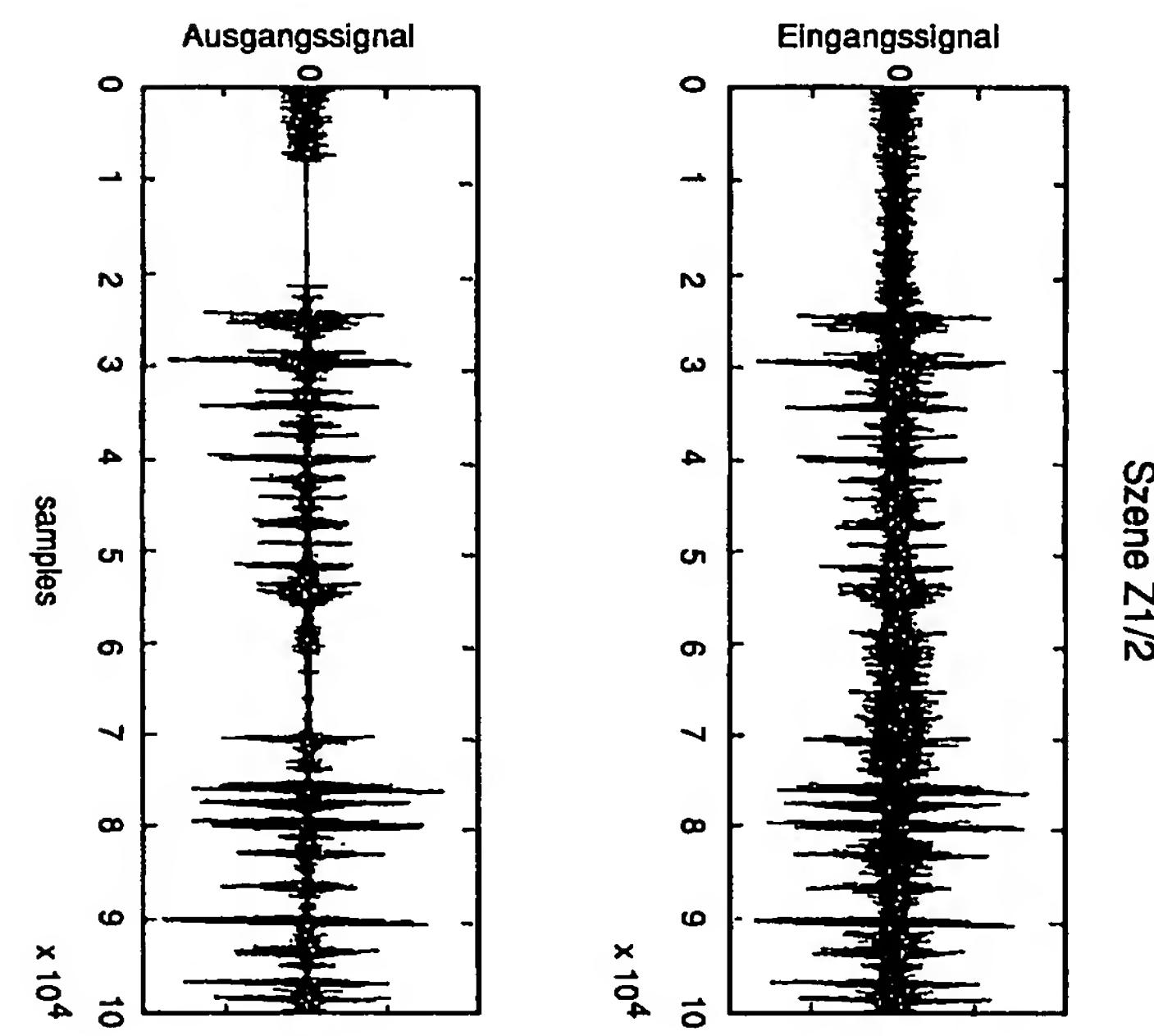


FIG.13

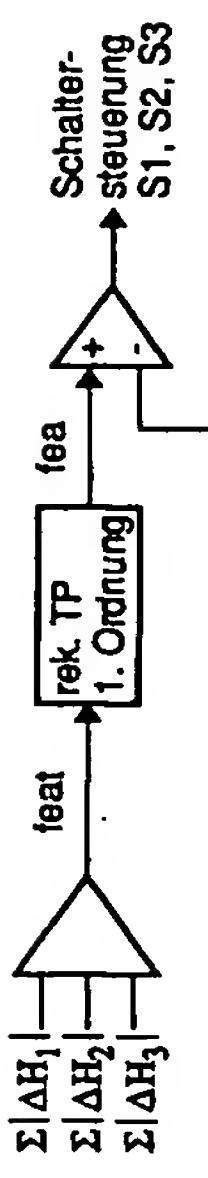


FIG.8a

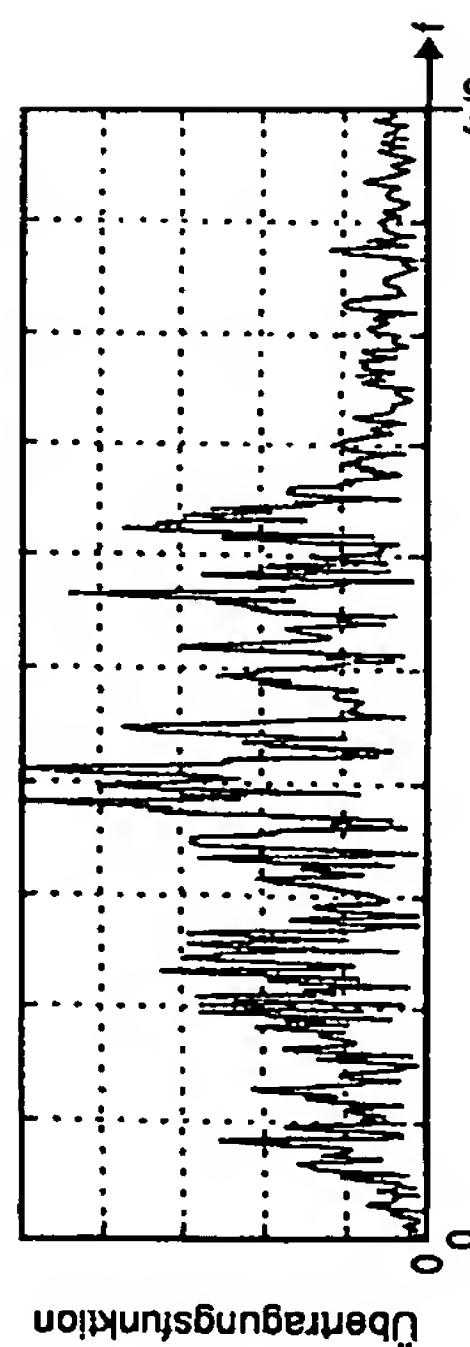
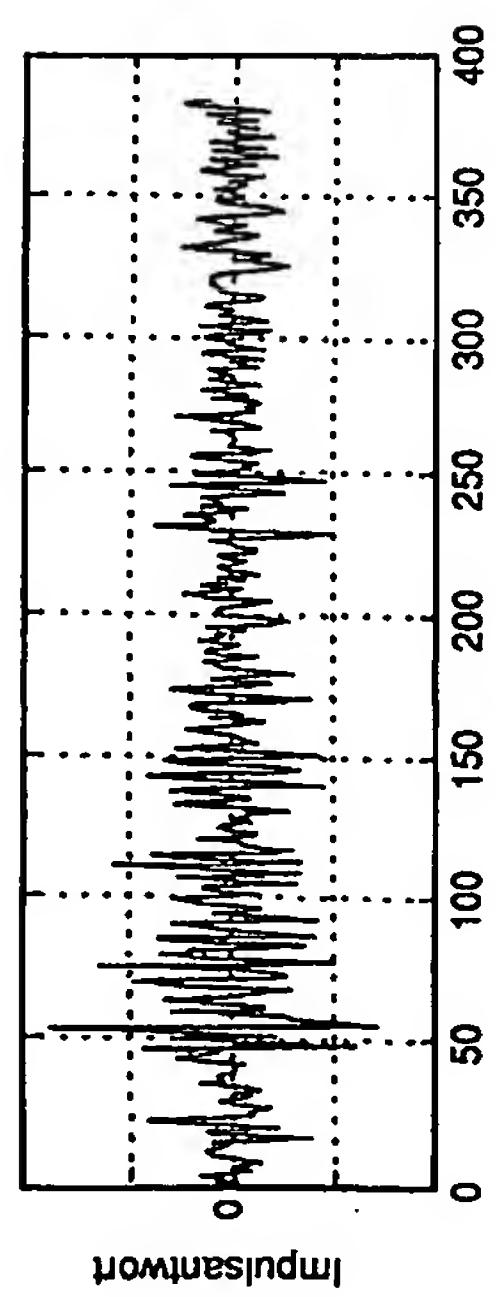
Szene Z1/1
Impulsantwort

FIG.12

Szene Z0

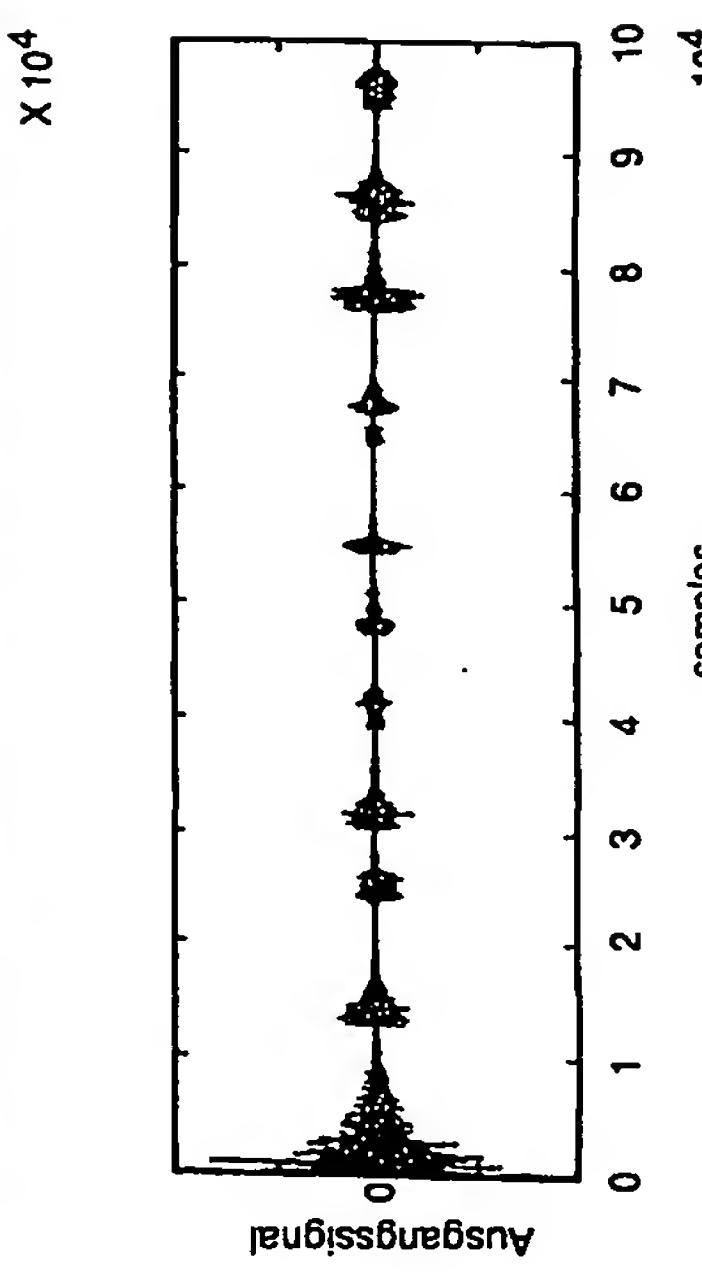
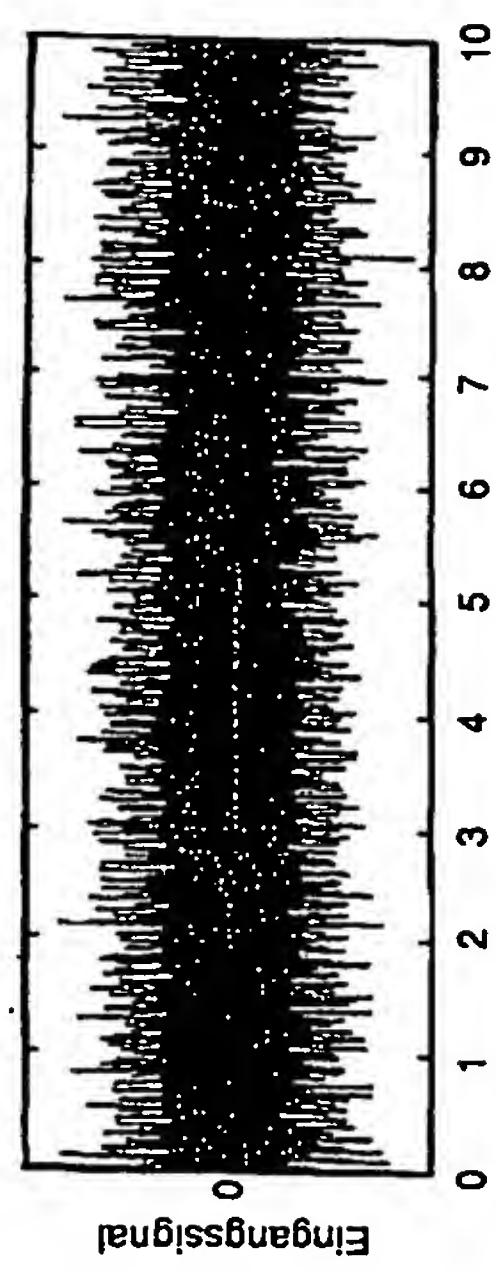


FIG.9

Szene Z0

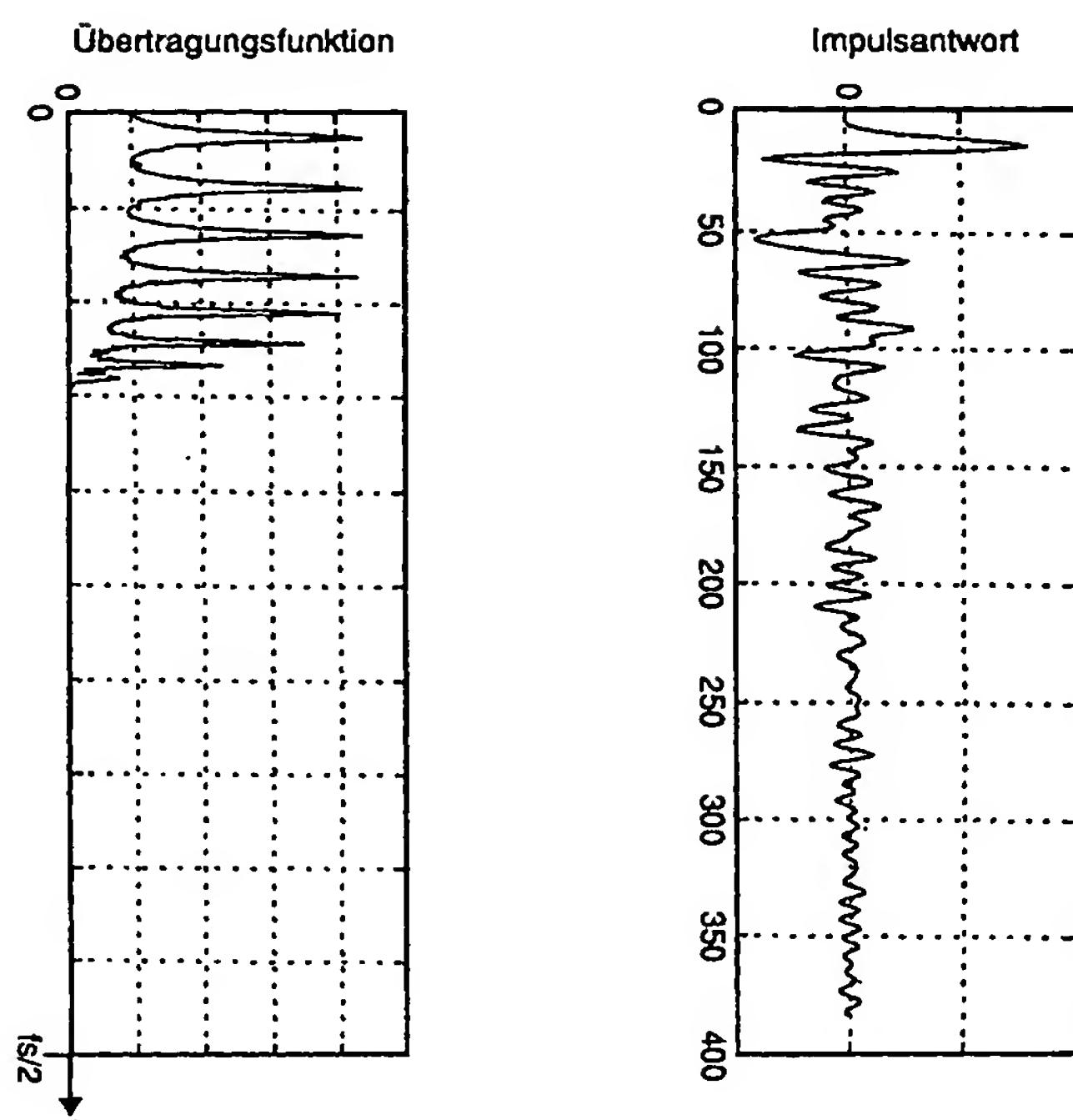


FIG.10

Szene Z1/1

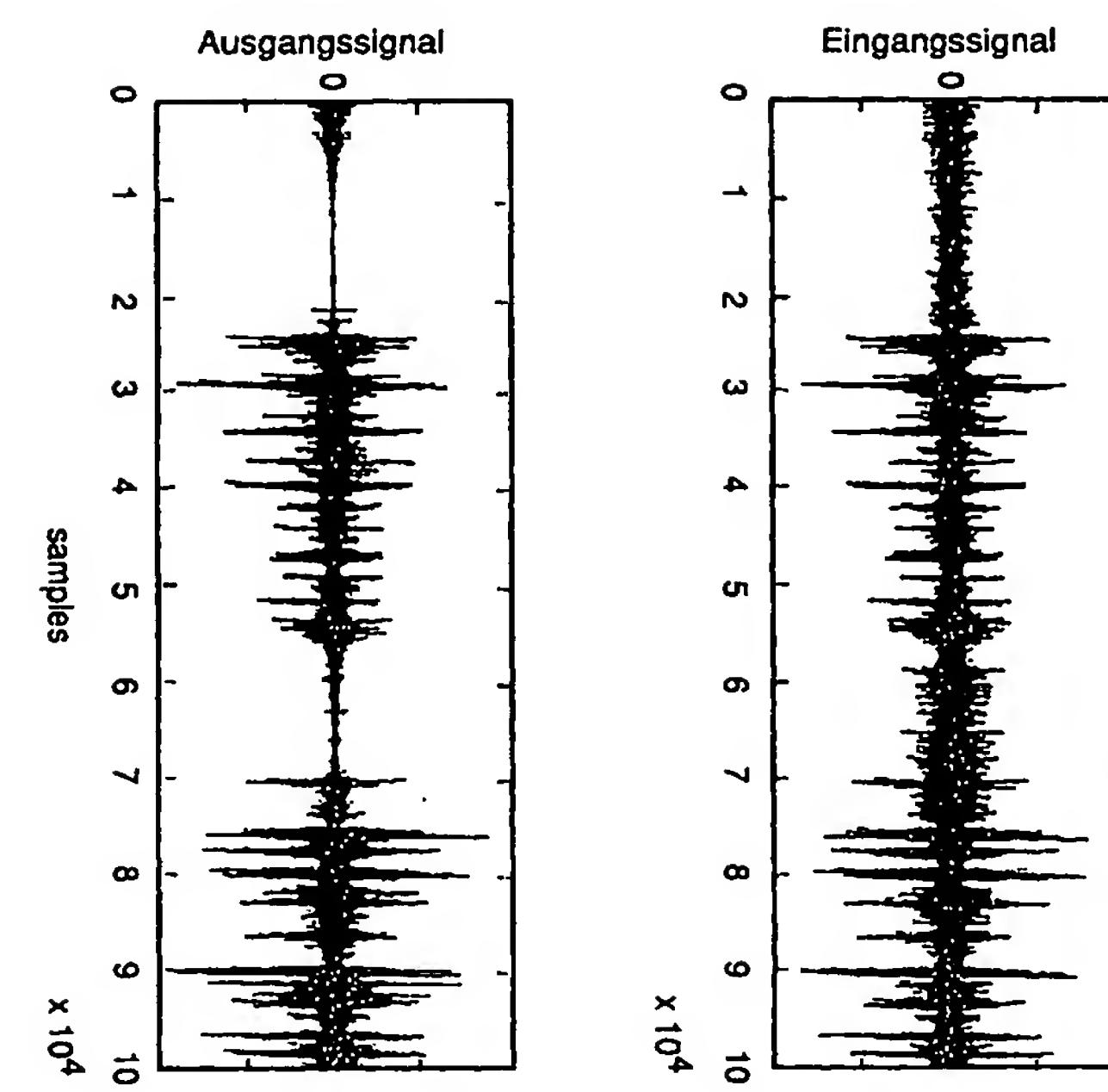


FIG.11

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT OR DRAWING
- BLURED OR ILLEGIBLE TEXT OR DRAWING
- SKEWED/SLANTED IMAGES
- COLORED OR BLACK AND WHITE PHOTOGRAPHS
- GRAY SCALE DOCUMENTS
- LINES OR MARKS ON ORIGINAL DOCUMENT
- REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox

This Page Blank (uspto)